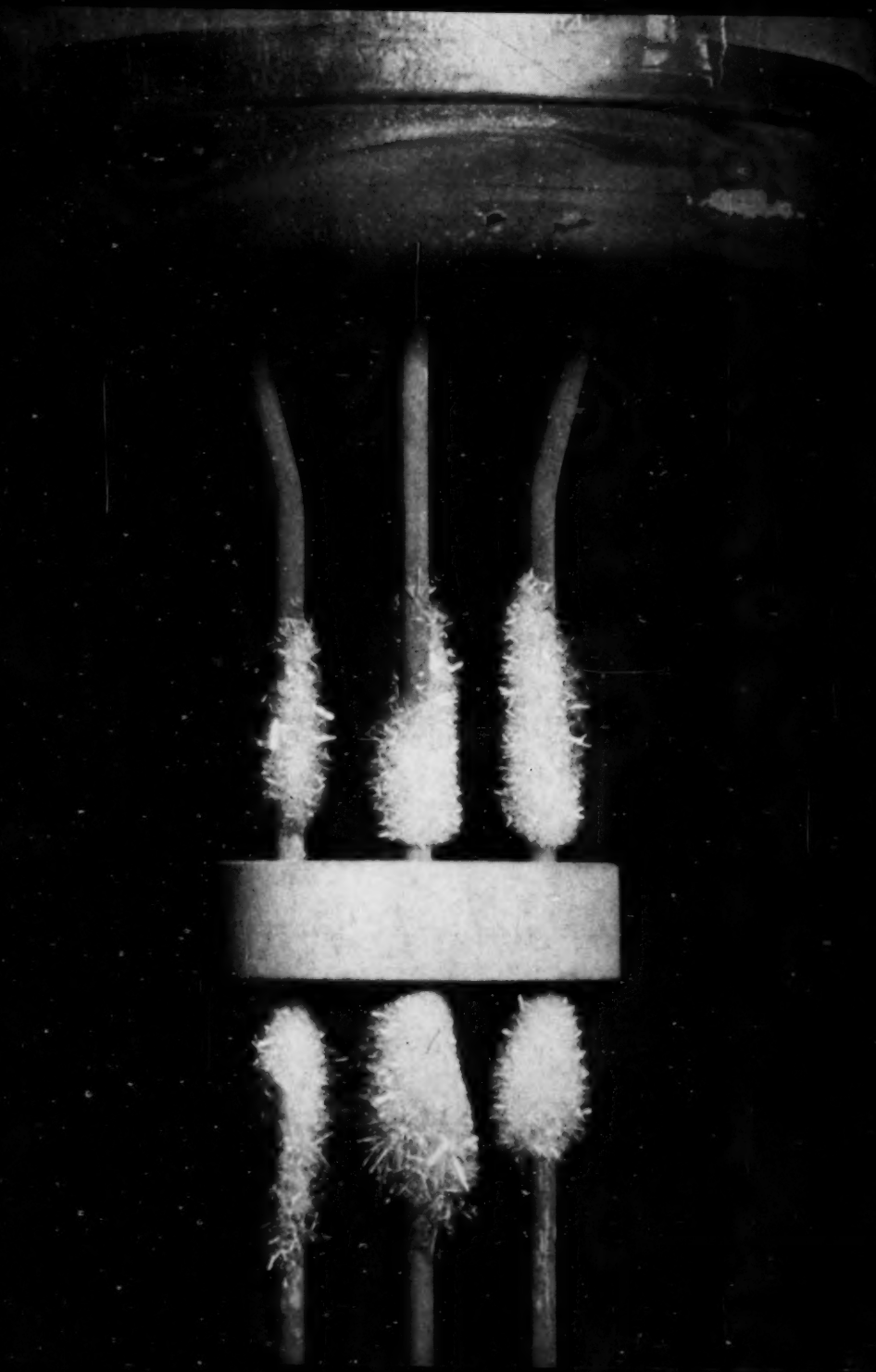
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 17, No. 11 NOVEMBER, 1946

In This Issue

Theoretical Physics in the Metallurgical Laboratory of Chicago E. P. WIGNER 857

Contributed Original Research

Composition and Diffraction Effects in X-Ray Microradiographs R. SMOLUCHOWSKI, C. M. LUCHT, AND J. M. HURD 864

Transient Analysis of Linear Systems, Using Underwater Explosion Waves M. F. M. OSBORNE AND J. L. CARTER 871

X-Ray Method for Measuring the Thickness of Thin Crystalline Films A. EISENSTEIN 874

New Bridge Photo-Cell Employing a Photo-Conductive Effect in Silicon. Some Properties of High Purity Silicon G. K. TEAL, J. R. FISHER, AND A. W. TREPTOW 879

Production of High Centrifugal Fields J. W. BEAMS, J. L. YOUNG, III, AND J. W. MOORE 886

Cathode-Ray Tube for Viewing Continuous Patterns J. B. JOHNSON 891

Improvements in the Stability of the FP-54 Electrometer Tube J. M. LAFFERTY AND K. H. KINGDON 894

Picturesque Oxidation of Molybdenum Vacuum-Tube Leads HERMAN A. LIEBHAFSKY 901

Luminescence of $(Zn, Be)_2SiO_4$:Mn and Other Manganese-Activated Phosphors JAMES H. SCHULMAN 902

Recent Ozone Investigations ARTHUR W. EWELL 908

Properties of Welded Contact Germanium Rectifiers H. Q. NORTH 912

Fine Structure of Viscose Rayon H. G. INGERSOLL 924

Theoretical Comparison of the Visual, Aural, and Meter Reception of Pulsed Signals in the Presence of Noise J. H. VAN VLECK AND DAVID MIDDLETON 940

Permeability of Different Rubbers to Gases and Its Relation to Diffusivity and Solubility G. J. VAN AMERONGEN 972

In Every Issue

Calendar of Meetings 900

Here and There 986

New Booklets 988

Index to Advertisers xxviii

BOARD OF EDITORS

ELMER HUTCHISSON, *Editor*

Associate Editors

R. BOWLING BARNES
OTTO BEECK
W. F. BUSSE
S. H. CALDWELL
CLARK GOODMAN
SELIG HECHT
W. JAMES LYONS
PAUL E. SABINE
FREDERICK SEITZ
J. R. TOWNSEND

Publication of the American Institute of Physics

ADMINISTRATIVE STAFF

HENRY A. BARTON, *Director*
MARGARET S. GRIFFIN, *Publications Manager*
THEODORE VORBURGER, *Business Manager*
JOHN T. TATE, *Adviser on Publications*

GOVERNING BOARD

PAUL E. KLOPSTEG, *Chairman*
WALLACE WATERFALL, *Secretary*
GEORGE B. PEGRAM, *Treasurer*
R. B. DOW
L. A. DUBRIDGE
W. F. FAIR, JR.
FLOYD A. FIRESTONE
HARVEY FLETCHER
R. C. GIBBS
K. S. GIBSON
W. H. MARKWOOD, JR.
A. H. PFUND
JOHN T. TATE
L. W. TAYLOR
A. G. WORTHING

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case School of Applied Science, Cleveland 6, Ohio.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U.S. AND CANADA	ELSEWHERE
To members of scientific societies	$5.00	$5.70
To all others	7.00	7.70

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 16, 1945—$101.50
Yearly back number rate: **$7.70**
Single copies: **$0.70** each.

Changes of address and proofs should be addressed to the Publications Manager.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1946, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 17, Number 11 November, 1946

Theoretical Physics in the Metallurgical Laboratory of Chicago*

BY E. P. WIGNER
Clinton Laboratories, Oak Ridge, Tennessee

DR. DARROW has suggested that a short introduction may enhance the usefulness of the more specialized papers to be presented to the Society by members of the old Theoretical Physics group of the Metallurgical Laboratory (Plutonium Project). Because the presentation of these papers will have to be very brief and because some of our collaborators could not present abstracts in time for the meetings, I was most happy to follow our Secretary's invitation to summarize our work in general terms.

There were, during the period extending from about the middle of 1942 until about the middle of 1945, that is for about three years, approximately twenty theoretical physicists assembled on the fourth floor of Eckart Hall who worked on those problems of chain reacting units which could be solved or at least attacked with the methods of theoretical physics. The membership of the group varied somewhat but I am happy to say, and do say it with a great deal of pride, that we formed a happy family and that I hardly remember a disagreement on non-technical points.

Most of our work was on very urgent problems and only a small fraction of it will bear publication in scientific periodicals. Our group had to do a great deal of engineering work, so much so that the calculation of liquid flow rates became to be considered to be part of our responsibilities. We calculated more than one I beam dimension, among similar matters, although we felt somewhat out of place when we did that. In addition, we had to maintain numerous contacts with the experimental groups and to take a lively interest in matters ranging from the fabrication and corrosion of aluminum tubes to the radioactivity induced in oxygen by neutron absorption. All this work was necessary and I do not hesitate to say in retrospect that our policy of assuming some of the functions which are usually reserved for engineers proved most useful. In the first place the Plutonium Project was not well provided with engineers in its early days—it had only two or three design engineers at the time when our plan for the *W* unit (which was later erected at Hanford) was virtually ready**; in the second place it was important, as in any new field such as that of chain reacting units was in 1942, that there be a few people at least who are sufficiently familiar with the whole picture to know about every difficulty and how it may be overcome. This second point will be very important in connection with Dr. Daniels' pile which is to be erected at Clinton. I fear that the present arrangement, which gives those who are

* An address presented to the American Physical Society at the Chicago Meeting, June 22, 1946.

** It may be interesting to remark that the date on the *W* report is January 9, 1943 and that it was issued just 42 days after the first nuclear chain reaction was experimentally established by E. Fermi and his collaborators (December 2, 1942). Of course, the relevant dimensions and sizes, etc., of the *W* pile had to be fixed much prior to that date, although, of course, they had to be verified experimentally later.

mainly interested in this pile even less responsibility than we had with respect to the *W* pile, will not help in meeting the time schedule which we read about in the newspapers. That time schedule was quite optimistic in the first place and I do not believe that the pile will be in operation before 1948.

As I said, knowing all details of the *W* plan was our main function and it took up most of our time. Practically all senior members of the group participated in it in one form or another although most of the burden was carried by Messrs. Friedman, Ohlinger, Weinberg, and Young and Miss Way. In addition, in the second year, Mr. J. A. Wheeler was transferred to Wilmington to give direct help to the DuPont Company. While work on the *W* plans was our most important function in the first two years, at the same time it was the one which is least suitable for being reported in public. In later years, work on the power production occupied most of our interest. However, this subject, to which Mr. G. Young has contributed so decisively, is still restricted from public discussion so that my report will deal, essentially, only with incidental studies which we have made. I shall deal with these under four headings:

1. Elementary theory of nuclear chain reactions.
2. More detailed theory of chain reactions.
3. Effect of radiation on matter.
4. Studies in theoretical physics.

I shall give an outline of our work in these fields in the above order.

1. ELEMENTARY THEORY OF NUCLEAR CHAIN REACTIONS

The great surprise about nuclear chain reactions was the ease with which they could be established. Szilard's paper of January 1940 already describes a workable arrangement. Our own early work in this field was not based on Szilard's paper but on Fermi's work, the concepts of which are less intricate than Szilard's. Ideas similar to Fermi's were developed also by others, notably by v. Halban; moreover, the whole work was duplicated, apparently without any major deviation, also by the German nuclear physicists.

In a chain reaction of the kind considered by us, uranium nuclei undergo fission and liberate neutrons. These neutrons are first fast but are soon slowed down by the moderator which is carbon (graphite) in our case. After being slowed down, the neutrons still diffuse around for a period of time before being absorbed. Most of them are absorbed by uranium, which then undergoes fission and emits the neutrons of the next generation. The ratio of the number of neutrons in one generation to the number of neutrons of the preceding generation is called multiplication constant and was usually denoted by k. Fermi's theory divides the problem of multiplication constants and critical sizes into two parts. The first problem is the calculation of the multiplication constant in an infinite medium, k_∞, which is usually referred to briefly as the multiplication constant. It depends only on the geometry and the materials of the chain reacting system and gives the ratio of the numbers of neutrons in successive neutron generations under the assumption that the same materials, arranged in the same geometry, extend all over infinite space.

The second problem is the calculation of a critical length which does not depend on the inner structure of the chain reacting unit but only on its size and shape. This critical length, or itsr eciprocal κ, permits one to calculate, from k_∞, the second kind of multiplication constant, k_{eff}. This gives the ratio of the numbers of neutrons in successive generations in a finite pile. This second kind of multiplication constant is, of course, the relevant one from a practical point of view. It depends not only on the materials and their arrangement into a lattice which determine k_∞, but depends also on the actual extension of the lattice, i.e., the size and shape of the pile. In a steadily running pile k_{eff} is always 1 and it exceeds 1 only when the power of the pile is increased, e.g., during startup, and then only very little. The k_{eff} is always smaller than k_∞ because in an actual, finite pile some of the neutrons of every generation diffuse out of the pile and do not contribute to the next generation. No such "leakage" exists in an infinite pile.

It would seem that only the effective multiplication constant has real significance but it turns out that the calculation of k_∞ is an almost necessary preliminary for the calculation of k_{eff}.

I will only sketch the calculation of k_∞ which

is already given, in principle, in the Smyth report. In order to calculate the number of neutrons of the next generation produced by one neutron of the present generation, one may start at the birth of one neutron. This occurs in the uranium lumps and the neutron has, originally, considerable velocity. As a result, it will be able to induce fission not only in the U^{235} nuclei but, what is more important because of their larger numbers, also in the U^{238} atoms. Competing with this process are the process of inelastic scattering by U atoms by which the original neutrons may be slowed to a velocity below the fission threshold[1] of U^{238}, and the process of escape of the neutrons from the U lump into the moderator.

The importance of fast fission was recognized by Szilard and his collaborators. The rest of the factors making up k_∞ were all recognized before, and are contained also in Fermi's considerations.

Let us assume that the original neutron generates $\epsilon - 1$ further neutrons by fast fission. As Fermi has mentioned in his address to this meeting $\epsilon - 1 \approx .03$ so that we arrive with $\epsilon \approx 1.03$ neutrons just below the fission threshold. Most of these neutrons diffuse out into the moderator and are slowed down to thermal energies. Some of them occasionally enter the uranium and are absorbed there by one of the numerous resonance levels of the U^{238}. These absorptions do not lead to fission and constitute an actual loss of neutrons. The importance of this process has been already recognized by N. Bohr in 1940, and others. Only when the neutron has lost sufficient energy to be below the lowest resonance level of U^{238}—which is, according to data in the literature[2] at about 5 ev—is it safe from this fate. The probability that a neutron will escape resonance capture is usually denoted by p. It is a number smaller than 1. As a result of the resonance absorption, we arrive with ϵp neutrons below the energy of 5 ev instead of the ϵ neutrons which we had just below the fast fission threshold.

It would lead too far to describe the actual calculation of p. Among all the processes which contribute to the chain reaction, the resonance absorption is the only one which was not really understood when we started our work. S. M. Dancoff and I were the ones who were most interested in the physical principles which determine the resonance absorption of macroscopic bodies, but ideas similar to ours were developed also by others. The actual calculation of p was described by R. F. Christy, A. M. Weinberg, and myself, although many others, including H. L. Anderson, contributed to it. The material constants necessary for the calculation were measured by Creutz, Jupnik, Snyder, and R. R. Wilson in Princeton, and later by Mitchell's group at the University of Indiana.

We now have ϵp neutrons with an energy below the resonance levels of uranium. According to theory, they will be slowed down to thermal energies by the moderator. After that, they will be absorbed, some of them by the moderator and the impurities present in the pile, some of them by the uranium. Fermi denotes this last fraction by f so that, altogether, ϵpf thermal neutrons are absorbed by the uranium, giving

$$k_\infty = \epsilon p f \eta \quad , \tag{1}$$

secondary neutrons, η being the number of fission (fast) neutrons produced in the uranium per thermal neutron absorbed. The principles for the calculation of the "thermal utilization" f were established independently by Fermi, Placzek, and our group. The formulae which we used were derived by Christy, Mrs. Monk, Plass, and myself in a way which is similar to the calculation of wave functions in metals by the cellular method.[3]

On the whole, the calculation of the multiplication constant for an infinite lattice is quite "straightforward" and one of the great surprises of the Plutonium Project was how easy it was. Mr. G. N. Plass and myself happened to be the ones who attempted to calculate the "optimal lattice" (i.e., the lattice with the highest k_∞) early in 1942. Although the physical constants were not known at that time too accurately, the dimensions we obtained (later incorporated in the first chain reacting unit) are now believed to give a k_∞ just $\frac{1}{2}$ percent short of the k_∞ of the real optimal lattice. We are quite convinced that

[1] Haxby, Shoupp, Stephens, and Wells, Phys. Rev. **57**, 1088A (1940); **58**, 199A (1940).
[2] H. L. Anderson, Phys. Rev. **57**, 566 (1940).
[3] Cf. e.g., F. Seitz, *The Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), Chap. IX.

any reasonably competent people would have arrived at the same results. In later years, computations of k_∞ were much facilitated by diagrams prepared for this purpose by Mrs. Monk and Mrs. Uchiyamada, under Professor Wheeler's direction. The calculation of k_∞ was extended to all sorts of lattices, containing heavy and ordinary water, etc. Most of this work was done by A. M. Weinberg and his collaborators, Mrs. Monk, Mr. Plass, Mrs. Uchiyamada, Mr. Stephenson and others. Qualitatively, the results were quite similar for all systems considered.

In spite of this, the properties which make a lattice optimal are not very simple. One may note that it is good if the high energy neutrons remain in the uranium to give a high ϵ. On the other hand, it is best if the lower energy (resonance) neutrons keep out of the uranium as much as possible so that p may remain reasonably close to 1. Again, the thermal neutrons should return to the uranium to give a high f—as close to 1 as possible. These conflicting requirements determine the geometry of the optimal lattice, i.e., give the ratio of the amounts of moderator and uranium as well as the lattice constant. However, even relatively large deviations from the optimal dimensions do not decrease k_∞ to a very great extent.

The foregoing describes the calculation of k_∞. Although Fermi has given a method for calculating k_{eff} from k_∞, I will not give his method here but will turn to the more advanced theory which permits a direct calculation if k_{eff}.

2. MORE DETAILED THEORY OF CHAIN REACTIONS

The more detailed theory of chain reactions should provide more accurate methods both for the calculation of k_∞ and also for the calculation of k_{eff}. However, as far as k_∞ is concerned, only few improvements were made. None of these improvements occurred in the calculation of ϵ and p, only one occurred in the calculation of f.

The behavior of "thermal" neutrons in a moderator-uranium lattice is far from simple. Evidently, it would take infinitely many collisions to establish real thermal equilibrium between the neutrons and the moderator, and in a well-designed lattice the neutrons will be absorbed by the uranium after a relatively small number of collisions. As a result, the energy spectrum of the neutrons will remain quite complex and their average energy will stay considerably above $\frac{3}{2}kT$. This average energy will be different even at different points of the lattice. The actual energy distribution will be influenced by the absorbing power of the material as well as by its moderating power. The latter is influenced in turn by the atomic weight of the moderator, by Fermi's chemical binding effect and by the crystalline nature of the moderator which gives a considerable anisotropy to the scattered (refracted) neutrons.

The only serious attempt to take these factors (excepting the crystal effect) into account is due to E. Teller and his collaborators, mainly N. Metropolis and P. Morrison. A more rigorous but much more formal attempt later by E. J. Wilkins and myself did not contribute much to the qualitative picture. Wick reported on Thursday about some work which he did on this problem. Teller's work gave, at least, an approximate measure for the difference in the effective temperature of the neutrons and the moderator. In spite of this we are far from having an adequate knowledge of the energy spectrum of the neutrons in a chain reacting unit.

Moreover, the problem of calculating f remains far from being simple even if the energy spectrum of the neutrons is known. It is, in fact, quite complicated even if one assumes that all the neutrons have the same energy. The reason for this is that the ordinary diffusion theory proves to be quite inadequate. G. Placzek carried out the most accurate calculations for the diffusion of monoenergetic neutrons. Some of his results were obtained also by German and Italian theorists[4] and published. Our work along this line was not pushed with much vigor because we were, perhaps, too well aware of the inadequacy of the model which uses monoenergetic neutrons. Actually, there is evidence that the errors in our primitive diffusion equations are quite substantial and in the direction indicated by Placzek's work.

There is no relevant difference between the fast effect in a finite and infinite lattice. However, the probability p that a neutron with an energy

[4] G. C. Wick, Zeits. f. Physik **121**, 702 (1943).

just below fission threshold should become a thermal neutron is smaller in a finite lattice than in an infinite one because in addition to being captured by the uranium, some neutrons will be lost from a finite lattice by "leaking" (diffusing) out of it. This leakage was calculated by Fermi and his co-workers[5] even before fission was discovered. For a finite lattice, their work gives

$$p_{\text{eff}} = p \exp(-\tau\kappa^2), \tag{2}$$

where τ is one-sixth the mean square distance, in an infinite lattice, between the point where the neutron originated and the point where it becomes thermal. The quantity κ^2 is the ratio $-\Delta n/n$ where n is the average of the neutron density over a lattice cell and will be discussed presently.

According to (2) the effective p is smaller than it would be for the constant n of an infinite lattice, i.e., for $\kappa = 0$. The leakage depends on the "age" τ, which on its turn increases with increasing mean free path of the neutrons in the moderator and with the number of collisions which are necessary to slow down the neutrons to thermal energies. The quantity τ, and hence the leakage, is smallest in a water moderated pile and much greater in a graphite moderated pile.

Just as the fraction of neutrons which are slowed down to thermal energies in the pile is, because of the leakage, smaller in a finite than in an infinite pile, so is the fraction of thermal neutrons absorbed by the uranium decreased by the escape of some of the thermal neutrons from the pile. The equation analogous to (2) is

$$f_{\text{eff}} = f(1+L_p^2\kappa^2)^{-1}. \tag{3}$$

The significance of κ in (3) is the same as in (2), that of L_p^2 similar to that of τ in (2): L_p^2 is one-sixth of the mean square distance in an infinite lattice between the point where the neutron becomes thermal to the point to which it has diffused when it gets absorbed. L_p is also called the diffusion length of thermal neutrons in the lattice because the n decreases with an exponential relaxation distance L_p

$$n \sim \exp(-x/L_p) \tag{4}$$

in a region in which no thermal neutrons are produced. G. N. Plass showed, by means of a calculation which is similar to Bardeen's work[6] on metallic wave functions, that

$$L_p^2 = L_m^2(1-f) \tag{5}$$

is a very good approximation for L_p if L_m is the diffusion length in the pure moderator, without uranium lumps.

The condition that a lattice can maintain a chain reaction at a steady rate is that $k_{\text{eff}} = 1$, i.e., that

$$k_{\text{eff}} = \epsilon p_{\text{eff}} f_{\text{eff}} \eta = 1. \tag{6}$$

Using the expressions (2) and (3), this becomes

$$\epsilon p f \eta \exp(-\tau\kappa^2)(1+L_p^2\kappa^2)^{-1} = 1$$

or, by (1)

$$k_\infty = (1+L_p^2\kappa^2)\exp(\tau\kappa^2), \tag{7}$$

an equation essentially identical to one already obtained by Fermi.

This last equation can be considered to be an equation for κ which, in its turn, will be seen to depend only on the size and shape of the pile. Hence (7) gives us the size of a pile if its shape and internal structure, in particular its infinite multiplication constant k_∞, are given.

The connection between the quantity κ and the size and shape of the pile is established by the classical equation

$$\Delta n + \kappa^2 n = 0 \tag{8}$$

in which the average neutron density n is subject to the boundary condition that it vanish at the outer boundaries of the pile. It is well known that (8) allows a solution only for definite, discreet values of κ^2 which depend on the size and shape of the region on the boundary of which n has to vanish, i.e., on the size and shape of the pile. Only for the smallest of these κ^2 is n positive throughout and this smallest κ^2 is the one which occurs in (7). Equation (8), so to say, gives an effective dimension κ^{-1} to every size and shape and (6) shows how this effective dimension affects the effective multiplication constant. If the κ of the pile, as defined by (8), is larger than the solution of (7), the pile is under critical, its effective multiplication constant smaller than 1. If the solution of (7) is larger than the κ satisfying (8), the pile is above critical.

[5] E. Fermi and F. Rasetti, Ricerca Scient. 9, 472 (1938); G. Placzek, Phys. Rev. 69, 423 (1946).

[6] J. Bardeen, Phys. Rev. 49, 653 (1936). Cf. also reference 3.

The quantity n to which (8) applies is the average neutron density, the average to be taken over a cell. Evidently, an equation, applying to such an average as (8) does, can be accurate only if this average does not change too rapidly from cell to cell. The relation of the n of (8) to the actual neutron density is similar to the relation of the macroscopic density of bodies to their rapidly fluctuating density as given by their atomistic structure. The theory of Eq. (8) is therefore called the macroscopic pile theory while the quantities of Eqs. (1) to (7) are concepts of the microscopic pile theory. Actually, (8) is only the simplest equation of macroscopic pile theory, which applies if the spatial variation of the neutron density is independent of energy. This is an important particular case but does not hold in general. For instance, most control rods absorb only low energy, thermal neutrons. The surface of a control rod is, therefore, a boundary where the density of thermal neutrons vanishes. However, the density of fast neutrons does not vanish at the surface of the control rod and the densities of fast and of slow neutrons are not proportional any more. Problems of this nature call for more complicated equations than (8). The most important results toward the solutions of these problems are due to Messrs. F. L. Friedman, A. M. Weinberg, and J. A. Wheeler.

Even the simple Eq. (8) raises a number of interesting problems. If the shape of the pile is at all complicated—which is almost invariably the case if the chain reacting material is liquid—the solution of (8) could be obtained only by perturbation methods. Some of these show a remarkable similarity to the Rayleigh-Schrodinger method with which we are familiar from its application to quantum-mechanical problems. We owe many interesting results on (8) to Messrs. F. Murray, L. W. Nordheim, and H. Soodak.

A good part of the work in this connection is too special to be taken up in detail. Another part will be dealt with by the speakers following my address so that I may close the subject of the calculation of multiplication constants and critical sizes. I would not like to do this before emphasizing that, in my opinion, a good deal of work remains to be done in this field. In particular, the behavior of "thermal" neutrons in the pile and the transition from fast to thermal energies requires further clarification both from the experimental and from the theoretical side. But there remain interesting details to be worked out in almost any part of the theory. There are, also, some problems which have already commanded considerable attention but which I have not even touched. Chief among these is the change in the neutron densities with time if Eqs. (6), (7) are not exactly fulfilled and the pile is either below or above critical. Messrs. R. F. Christy, L. W. Nordheim, and J. E. Wilkins were particularly active in this field.

3. EFFECT OF RADIATION ON MATTER

The radiation densities, both γ and neutron, are higher in a plutonium producing pile than can be maintained outside the pile for extended periods. The effect of these radiations on the structure of materials was one of our early concerns from the theoretical point of view. The experimental work was carried out in the Chemistry Division. Dr. M. Burton reported at the Atlantic City meeting of the American Chemical Society about his, his collaborators' and Dr. J. Franck's work on the subject. On the theoretical side, M. Goldberger, R. S. Mulliken, and F. Seitz shared my interest in the subject which still has some aspects about which we cannot talk freely.

Clearly, the collision of neutrons with the atoms of any substance placed into the pile will cause displacements of these atoms. If the substance is a chemical compound, the displacement will result in chemical changes which were, of course, investigated already before chain reacting units came into being and are summarized, e.g., in the booklet of Lind.[7] All these changes are much more intense in the pile owing to the more intense radiation. But substantial effects can be expected in elementary substances also. The matter has great scientific interest because pile irradiation should permit the artificial formation of displacements in definite numbers and a study of the effect of these on thermal and electrical conductivity, tensile strength, ductility, etc. as demanded by theory. One may expect that

[7] Cf. e.g. S. C. Lind, *Chemical Effects of α Particles and Electrons* (Chemical Catalogue Company, New York, 1928).

studies of the solid state, particularly of the structure sensitive properties, will be greatly stimulated by the additional experimental facility given by the pile.

Before a final interpretation of the experimental results can be made, our knowledge of the ranges of low energy ions will have to be extended. It is on this subject that most of Messrs. Goldberger's and Seitz's work was concentrated and you will hear from them presently. A good deal of the rest of our work was speculation which will either be confirmed or refuted by future experiments.

4. THEORETICAL PHYSICS

As I emphasized before, real theoretical physics was always somewhat of a stepchild. This situation hardly could be remedied in view of our many pressing obligations. We tried to free Mr. S. M. Dancoff as much as possible from the pressure of urgent work and he carried out several investigations which are of considerable general interest. There is only one other function which we did not permit to be pushed into the background and this was the problem of keeping tables of nuclear constants up to date. Mrs. Uchiyamada was in charge of these tables but she received a good deal of help from the Project Information Department, in particular from Mr. Goldsmith who made a really brilliant contribution. We were happy to learn that these tables may be at least partially published soon.

The theoretical work of the group fell into two categories; help with the evaluation and planning of experimental work, and real theoretical work. Into the first category falls the work of Messrs. Cahn, Schweinler, Weinberg, and others on the so-called pile oscillator. This is an instrument which permits an absorber of known or unknown neutron absorbing characteristics to be put into periodic motion in the pile. The oscillation of the neutron absorber causes intensity waves to spread all over the pile. These waves are similar to the temperature waves in the earth, generated by the daily and yearly heat fluctuations of the heat input on the earth's surface. The amplitude and wave-length of the waves permits one to evaluate the characteristics of the neutron absorbing oscillators and of properties of the pile.

The work on neutron diffraction received considerable attention on the part of Goldberger and Seitz. They interpreted and extended Weinstock's results[8] considerably and took into account phenomena not previously considered. Their work is being continued by Mr. M. Moshinsky in Princeton. There is little need for my going into details on this question since Seitz and Goldberger will tell you about them presently.

As a last example, I would like to mention Dancoff's work on short range α-particles. This work actually started because of some acute problem which was practically forgotten by the time Dancoff took over. He noticed that the intensity of short range α's is often anomalous in the light of Gamow's theory[9] which stipulates that the α-particle is emitted by an excited residual nucleus. Dancoff investigated several other mechanisms among which the excitation of the residual nucleus by the α-particle after it has already penetrated the potential barrier seems to be the most important. These theoretical investigations have now received added interest in view of Chang's[10] experimental results. Chang has discovered Dancoff's mechanism independently. Dancoff will tell you about this phase of his work in the afternoon.

I would not like to close my review of the work of our group without expressing my sincere thanks to all members of the group for their most unselfish and loyal cooperation. I have to extend my apologies to those whose work I may have slighted. As I said before, we had a very concrete objective in mind during the course of the work and the most important problems solved were not always the ones which now appear most worth remembering. We have, however, encountered a good many interesting problems, several of which will bear a great deal of further study. A glimpse at these problems, together with the knowledge of the importance of our aims, contributed a great deal to making the relations between our group cordial.

[8] R. Weinstock, Phys. Rev. **65**, 1 (1944).
[9] Cf. G. Gamow, *Structure of Atomic Nuclei* (Clarendon Press, Oxford 1937), p. 104 ff.
[10] W. Y. Chang, Phys. Rev. **69**, 60 (1946).

Contributed Original Research

Composition and Diffraction Effects in X-Ray Microradiographs

R. SMOLUCHOWSKI,* C. M. LUCHT, AND J. M. HURD
General Electric Company, Schenectady, New York
(Received June 1, 1946)

MICRORADIOGRAPHY is not a new art. However, its systematic application to various problems is rather recent. The principal idea of the method is extremely simple: A thin sheet of a material is placed on a photographic plate and exposed to x-rays. The resulting image on the photographic plate is enlarged or examined under a microscope; the art consists in obtaining reproducible results, which can be unambiguously interpreted. The early work in the field of microradiography was primarily in the field of biology[1] and geology.[2] The first to use it for the study of metals were Fournier[3] and later Trillat.[4] The latter author especially has obtained interesting examples of the distribution of various elements in alloys as well as of internal defects. Since then a number of authors have improved and described various aspects of this method.[5] The purpose of the present investigation[6] was not so much to improve the technique as to make a careful comparison with results obtainable with a conventional light microscope and to look for and interpret new *diffraction* effects.

THEORETICAL

The theory of concentration effects is very simple and has been discussed in numerous places.[5] Only the fundamental facts will be recalled here: Every chemical element has its characteristic x-ray absorption spectrum. Thus by a proper choice of x-ray wave-length (anode material and voltage) one can produce a difference in absorption of various constituents of an alloy. Concentration gradients in the direction perpendicular to the x-ray beam show up as gradients in blackening. A uniform thickness of the metallic foil, smaller than the size of the concentration non-uniformities, is naturally assumed. Apart from a careful consideration of the wave-lengths and absorption coefficients, two facts should be borne in mind in the interpretation of the results; First, that a microradiograph is the projection of the whole thickness of the foil on the photographic plate and not only of its surface and, secondly, that except in the proximity of an x-ray absorption edge, the absorption does not depend upon the chemical state of the atom.

The theory of diffraction effects is equally simple. However, it was not described previously in this connection and so a short description may not be out of place. Let us assume that the shaded elongated areas in Fig. 1 indicate cross sections through various grains in a metallic foil placed perpendicularly to the plane of the drawing, and that the lines themselves indicate orientation of a set of crystallographic planes in each grain. Monochromatic x-rays passing through the foil in the direction indicated by the arrows will undergo absorption according to the usual

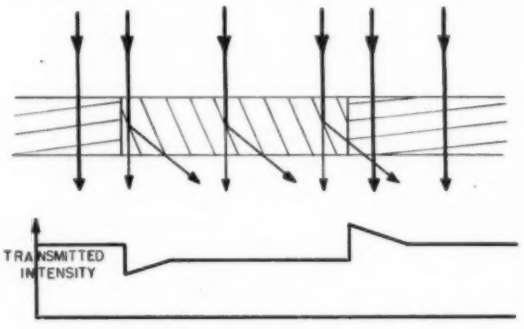


FIG. 1. Diffraction effects on grain boundaries.

* Now at the Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

[1] A. Dauvillier, Comptes rendus **190**, 1287 (1930).
[2] P. Goby, Comptes rendus **156**, 686 (1913).
[3] F. Fournier, Rev. Metallurgie **35**, 349 (1938).
[4] J. J. Trillat, Rev. Scientifique **78**, 212 (1940).
[5] For instance: G. L. Clark and W. M. Shafer, Trans. A.S.M. **29**, 732 (1941). S. T. Gross and G. L. Clark, Ind. Radiography **1**, 21 (1942). S. E. Maddigan, J. App. Phys. **15**, 43 (1944). S. E. Maddigan and B. R. Zimmerman, Trans. A.I.M.M.E. **156**, 33 (1944).
[6] R. Smoluchowski, C. M. Lucht, and J. M. Hurd, Phys. Rev. **68**, 100 (1945).

law relating loss in intensity to the thickness of the traversed material. However, if the direction of these x-rays happens to form a Bragg angle α with a set of planes in a particular grain, then the radiation will be diffracted and it will come out from the foil at an angle 2α with the original direction. Naturally the path of these x-rays in the foil is longer than if they were not diffracted, and so the intensity transmitted through that grain will be correspondingly weaker. Because of the exponential character of absorption this difference may be quite large. This "effective thickness" may vary from grain to grain and so various degrees of transmission can be expected.

Grains adjoining that grain which happens to diffract the x-ray beam do not need to be themselves in an orientation suitable for diffraction. In general we would expect thus two effects to be observed at the grain boundary: If in one grain the x-ray beam is diffracted away from the boundary (in Fig. 1 the left boundary), then there should be a white line along the boundary. If, on the other hand, the radiation is diffracted towards the boundary, then in the adjoining non-diffracting grain we should observe a dark line along the boundary (right boundary in Fig. 1). If two adjoining grains diffract, then, depending upon the relative orientation of the grains, their boundary may have a light line on one side and a dark line on the other.

If the photographic plate is not in direct contact with the metal foil but at some distance, say 2 mm, from it, then one observes two areas for each diffracting grain: One is a white area where the x-ray beam would have reached the plate if it were not diffracted, the other is a black area where the diffracted beam reaches the plate. If the angle of diffraction was large or the absorption high, then the dark area may not be observable because not much intensity would emerge from the diffracting grain. The white and the dark area are essentially complimentary and can be readily correlated. Knowing the distance between the sample and the plate and the distance between the two areas the angle of diffraction and thus the index of the diffracting plane can be calculated. The white and dark lines along grain boundaries are limiting cases when the two areas, the white and the dark one, almost coincide on the photographic plate.

The condition for occurrence of a Bragg reflection is quite sharp and so even a small change of orientation of the metal foil with respect to the x-ray beam should produce a change in appearance of the microradiograph and in the relative intensity of the various grains. Since intensity of diffraction depends upon the index of the crystallographic plane which is oriented at the proper angle, it follows that various gradations of intensity will be present in a microradiograph of a non-oriented polycrystalline material. An oriented material will show a simultaneous "flashing" of many grains and so microradiography provides a new tool for a study of preferred orientation. If, instead of monochromatic (in practice, filtered) radiation, the full radiation of an x-ray tube is used, then the contrasts will naturally be lower.

In the discussion here it was assumed that the crystal lattice of the metallic foil is free from strain. Naturally this does not need to be true and the changes in microradiographs produced by strains turn out to be of great interest. Their study will be the subject of a later paper.

The two effects, composition gradients and diffraction, can be distinguished in a simple manner. It is sufficient to repeat the exposure with a slightly tilted sample. If the pattern remains the same, it is caused by concentration gradients, if it changes it is a diffraction effect. Usually a change of the angle of the order of a fraction of a degree is sufficient to distinguish between the two effects. Naturally a diffraction effect may be influenced by a concentration gradient. In that case by a proper choice of anode material one can reverse the concentration effect or eliminate the diffraction effect retaining the gradient of blackening due to concentration non-uniformities. Such a procedure may be necessary, for instance, in the case of a solid solution in which a steep concentration gradient produces a gradient of the lattice constant, and thus a gradient in the condition for diffraction.

EXPERIMENTAL TECHNIQUE

The main experimental difficulty in microradiography lies in preparing the metallic foils. However, with certain practice the necessary time is about the same as for preparing regular metallographic samples. A thin slice of metal

was cut off by a cut-off wheel and then ground and polished on both sides. During these operations, the thin foils were cemented to metal blocks so as to assure a perfectly flat surface. In certain cases, when it was advisable to avoid straining the sample during polishing operations, the electrolytic polishing technique was used. The sample was again cemented to a copper block taking care to establish electrical contact between the block and the sample.

The samples were usually about .05 mm (.002″) thick and were cemented over an appropriate hole in the center of a small brass disk. Figure 2 shows the construction of the camera in which the ring *R* with the sample *S* was placed in a centering depression and covered with a photographic plate. A cover *C* was fastened by means of two wing nuts.

The camera was placed within 12 cm from the focal spot of a General Electric X-Ray tube, Type CA5. The tube was mounted in such a manner that the x-ray beam emerged upwards in a vertical direction. If it was necessary to tilt the camera, then it was placed in a special support. By means of a fine dial indicator the angle of tilt could be measured to better than ½ a minute of arc. The photographic material was standard 548-O and 649-G H Eastman Spectroscopic Plates cut to appropriate size. Both types of plates were developed for 1½ minutes in D8 which we found to give smaller grain than the usually recommended D19. The photographic plates were examined with an ordinary microscope. If it was necessary for comparison to obtain a regular metallographic photograph of the foil, then it was etched in the usual manner. Inasmuch as etching produces surface irregu-

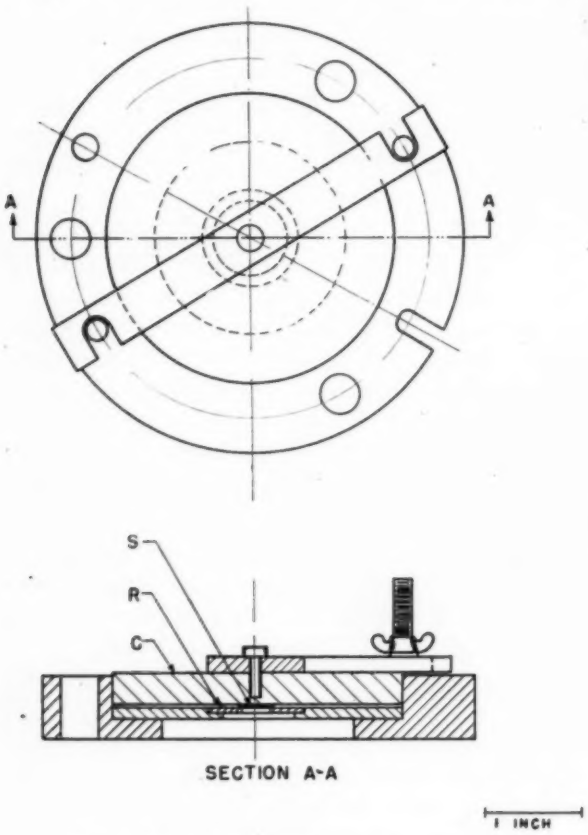


FIG. 2. Microradiographic camera. *S*—sample, *R*—supporting ring, *C*—cover.

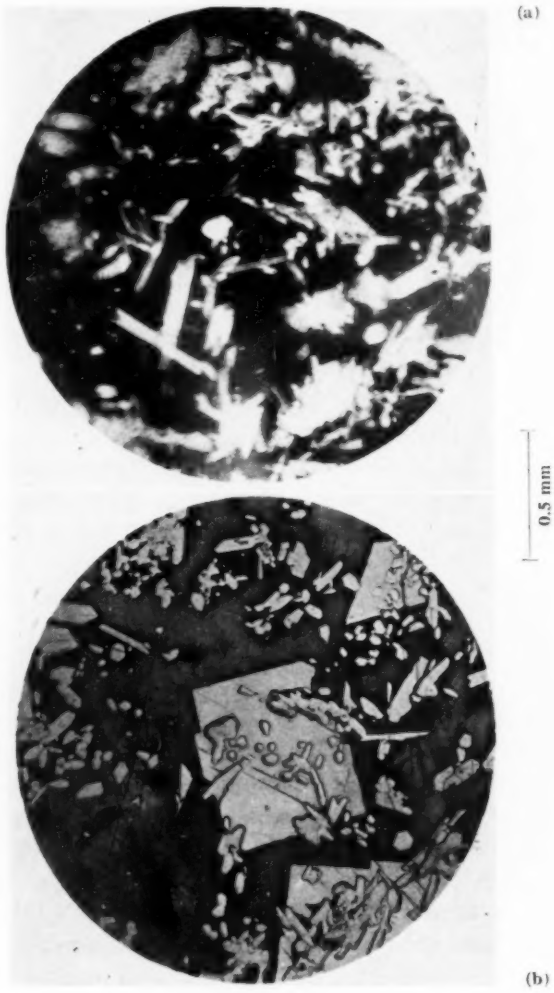


FIG. 3. Photo and x-ray microradiographs of a bearing alloy. The same area, 1.8 mm diameter. (a) x-ray, Cr anode, 20 kv. (b) photo.

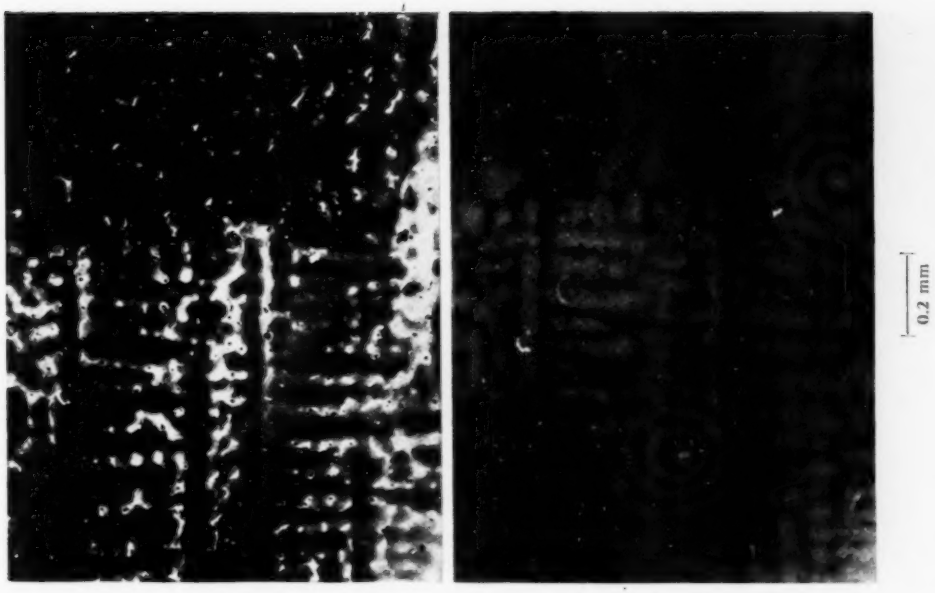


FIG. 4. Microradiographs of a cobalt alloy. The same area, 1.1×0.8 mm.
(a) Cr anode, filtered, 20 kv. (b) Mo anode, filtered, 20 kv.

larities and thus locally changes thickness of the foil, it is detrimental to a microradiographic study of other effects. For that reason all x-ray pictures were taken before the sample was etched because it was often impossible to remove the etching without destroying the thin sample. All illustrations shown in this paper are positives.

EXPERIMENTAL RESULTS

If the microradiographic technique would not show anything more than is readily visible under a light microscope, it would not have much significance. It is important, therefore, to stress the differences between the two methods.

A. Concentration Effects

1. Identification of phases

If there are several phases in a complicated alloy it is often very diffcult to estimate the distribution of the various constituents among the phases. Figure 3 shows a photomicrograph and an x-ray microradiograph of a babbitt alloy (82 percent Sn, 9 percent Sb, 9 percent Cu, and some minor impurities). The *area* is exactly the *same* in both pictures. The large square precipitates are much more transparent than the rest and so they are readily identifiable as containing primarily copper. Several small precipitates do not occur in the photomicrograph because they do not reach the surface. The light square in the lower right of the photomicrograph appears only very weakly in the microradiograph indicating that the precipitate is very shallow.

Another example of phase identification is provided in Fig. 4 which shows microradiographs of the same area of a cobalt alloy (64 percent Co, 26 percent Cr, 5 percent Mo, 2 percent Ni, 1 percent Fe, and impurities) made with filtered Cr and Mo radiations. Apart from the nicely visible dendritic structure, the finely divided phase is easily identifiable as a molybdenum rich phase. It is highly transparent to Mo radiation and very absorbing to Cr radiation. None of the other constituents would show a similar reversal. This can be seen from a comparison of the absorption coefficients per unit thickness for chromium and for molybdenum radiations. The respective values are: Co 1120, 374; Fe 910, 300; Mo 4500, 204; Ni 1290, 418; Cr 640, 213. The absorption coefficients of cobalt for both radiations are greater than the corresponding coefficients for iron and chromium and smaller than those for nickel. Only for molybdenum the first absorption coefficient is larger, the second smaller, than the corresponding values for cobalt.

In that connection it may be pointed out that

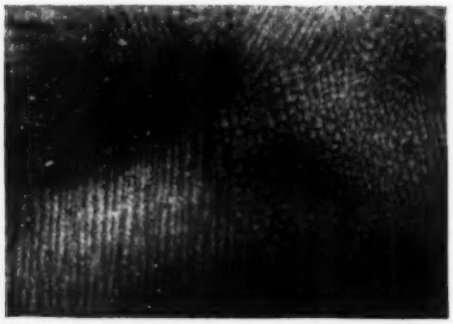


FIG. 5. Concentration gradients in a 50 percent Au, 50 percent Ag alloy. Mo anode, 35 kv. Area: 0.9×0.6 mm.

criteria for good contrast obtained on the basis of studies with unfiltered radiation are not well applicable if filtered radiation is used. Nevertheless, the following condition, suggested by Maddigan[5] is satisfied in the microradiographs here reproduced:

$$\exp |(\mu_1-\mu_2)t| < 0.9,$$

where μ_1 and μ_2 are absorption coefficients of the two neighboring areas and t the thickness of the layer.

2. *Concentration gradients*

If a metallic surface would etch uniformly, then a microscopic examination would not show any structure. In other words, what we see in such cases are only differences in etching. These differences are usually big enough to bring out boundaries between various phases in an alloy. However, if there is only a gradient of concentration, then it will show up only if there is a change in etching characteristics with concentration. This change, if any, occurs usually for a given critical concentration within the solid solution and so, instead of a gradient, there appears under the microscope an outline of the boundary between the regions of higher and of lower concentration than critical. Microradiographs are naturally very suitable for the study of concentration gradients. Figure 5 is a microradiograph of a 50:50 gold-silver alloy with very pronounced concentration non-uniformities which are due to rapid cooling during the process of crystallization.

Microradiographs also show concentration gradients surrounding separated phases. In Fig. 4 each molybdenum-rich phase is surrounded by a gradually vanishing gradient of concentration reaching deep into the surrounding matrix. Similar effect is visible in Fig. 6 in a 75 percent Mg, 25 percent Pb alloy, which is in "as cast" condition.*

Perhaps the most striking comparison of a microradiograph and a photomicrograph is in Fig. 7 which shows a 70 percent Al 30 percent Ag and a 90 percent Al, 10 percent Zn alloy, respectively. The two Al-Ag pictures show the same area and the contrast between the chaotic photograph and the highly organized x-ray picture is striking. Many details of the photograph appear to be accidental surface irregularities having no significance as far as the underlying, highly symmetric dendritic structures, as revealed by x-rays, is concerned. The white network is naturally silver. The Al-Zn alloy pictures do not show exactly the same area but a characteristic area from the same grain. (The area of the microradiograph was accidentally badly scratched after the x-ray pictures were taken.) Here the contrast between the large etched spots on the surface and the actually quite small precipitates surrounded by a zinc gradient is very striking. It is a typical example of etching

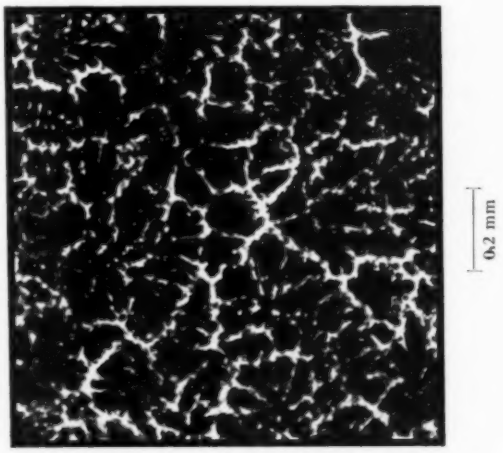


FIG. 6. Concentration gradients in a 75 percent Mg, 25 percent Pb alloy. Cu anode, filtered, 20 kv. Area: 1.1×1.1 mm.

* Solubility of Pb in Mg at 475°C is about 3.7 at. percent, at room temperature about 1.4 at. percent. The referee kindly pointed out that the blackening gradients may be caused by an oblique position of a grain boundary precipitate. Probably both factors play a role.

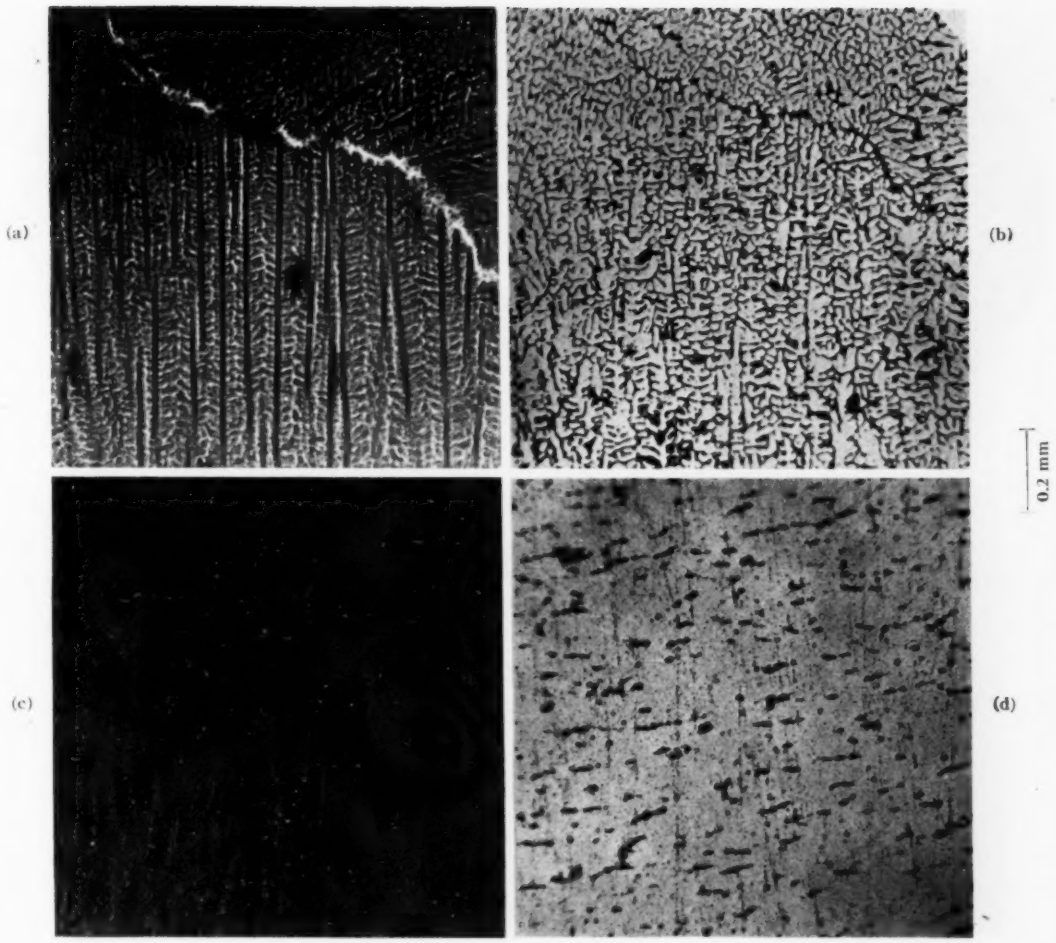


FIG. 7. Photo and x-ray microradiographs (area: 1.2×1.2 mm): (a) 70 percent Al, 30 percent Ag alloy, Mo anode, 35 kv. (b) 70 percent Al, 30 percent Ag alloy, photo. (c) 80 percent Al, 20 percent Zn alloy, Cr anode, 20 kv. (d) 80 percent Al, 20 percent Zn alloy, photo.

FIG. 8. Diffraction along grain boundary in an iron foil, Cr anode, 20 kv. Area: 0.8×0.9 mm.

of the "critical concentration" within the range of concentration gradient. These two examples show clearly how misleading conclusions can be drawn from conventional microscopic study of metallographic samples.

An interesting application of microradiography is the study of diffusion. One of the authors has used this method in the control of samples in a study of diffusion of copper in aluminum.[7]

B. Diffraction Effects

As discussed above, in microradiographs, there occur also diffraction effects connected with orientation of the individual grains. A few characteristic cases are illustrated in the following

[7] F. J. Radavich and R. Smoluchowski, Phys. Rev. **65**, 257 (1943).

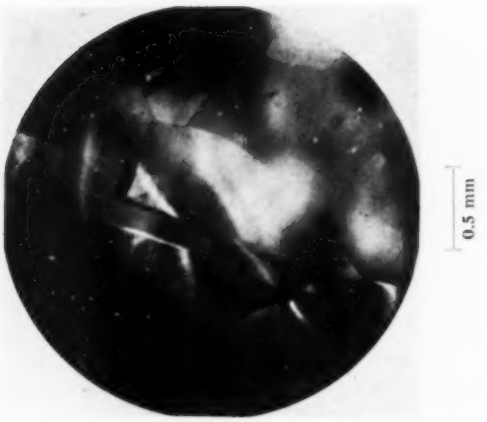


FIG. 9. Diffraction effects in a copper foil. Cu anode, filtered, 20 kv. Diameter 3.3 mm.

FIG. 10. Same as Fig. 9, tilted.

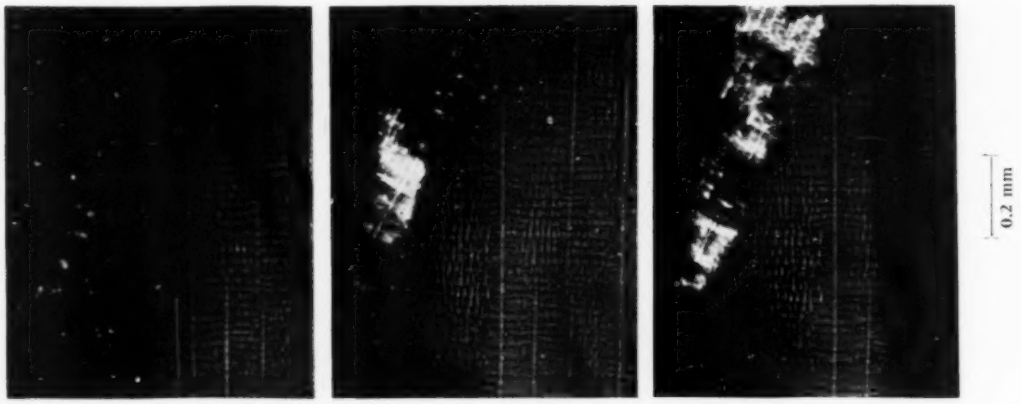


FIG. 11. Diffraction and concentration effects in a copper-beryllium alloy. Cu anode, filtered, 20 kv. Area: 0.9×0.7 mm.

microradiographs. Figure 8 shows a typical white line parallel to the grain boundary in pure iron. Naturally, the white line gradually disappears as the boundary turns into the direction of the diffraction. The diversity and clearness of the general diffraction pattern in a polycrystalline foil is illustrated in Fig. 9 which shows a microradiograph of an electrolytically polished copper foil. The differences in intensity of various grains and the relative displacement of the dark and the light areas are well visible. The small spots are caused by defects in the photographic emulsion. The influence of tilting on the general diffraction pattern is illustrated in Fig. 10 which should be compared with Fig. 9. The change in relative blackening is evident. Finally, Fig. 11 illustrates how a diffraction pattern can be differentiated from a concentration effect in the case of a copper-beryllium alloy. On tilting, the diffraction pattern due to some deformation changes while the dendritic structure remains unchanged. The diffraction from a single branch of a dendrite is particularly interesting, because it shows that even one individual member of a whole dendrite can be distorted. A more detailed microradiographic study of this alloy is under way.

These few examples indicate the correctness of the theoretical considerations and point to a new field of application of microradiography.

Transient Analysis of Linear Systems, Using Underwater Explosion Waves

M. F. M. OSBORNE* AND J. L. CARTER**
Naval Research Laboratory, Washington, D. C.

(Received June 7, 1946)

The indicial admittance of a simple mechanical system, a tourmaline crystal hydrophone, has been determined from its response to an underwater explosion wave. The indicial admittance is simply related to the observed response to the explosion wave—the observed response is just the indicial admittance "wrapped around" an exponential, $e^{-t/\tau}$. The relative frequency calibration of two hydrophones obtained by Fourier analysis of their responses to the explosion wave is in good agreement with the direct continuous wave comparison. The peak pressure of the explosion wave is also determined from this Fourier analysis. It is suggested that the explosion wave can be used as a tool in acoustical "circuit analysis" in much the same way that a square wave generator is used in electrical circuit analysis.

IN a previous paper[1] an iterative, numerical procedure was given for determining the indicial admittance, or response to the unit function, of any linear system, from its response to a known, arbitrary transient. In this paper an application of this method is made to a simple structure, a tourmaline crystal hydrophone. The transient used was the pressure wave from an underwater explosion—a No. 8 Dupont electrical blasting cap. This transient has a simple and known form,[2,3] that of an exponential decay, $p = p_{max}e^{-t/\tau}$. The time t is measured from the instant of arrival of the shock front, and τ is the "time constant" of decay.

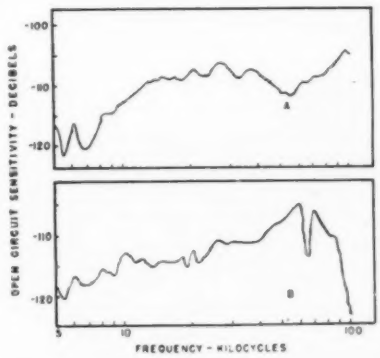


FIG. 1. Frequency calibration of hydrophones compared. 0 db = 1 volt/dyne-cm².

The explosion wave was also used to obtain the relative frequency calibration of two hydrophones, using methods quite similar to those used by Bedford and Fredenhall[4] in obtaining the frequency calibration of a television amplifier with a square wave generator. Since the absolute calibrations of the hydrophones were known independently, this analysis also gave information about the explosion wave.

The two hydrophones analyzed were similar and consisted of a stack of tourmaline disks, set in a metal sphere filled with castor oil, with a rubber acoustic "window" over the face of the crystals. The frequency response curves of the two hydrophones A and B are shown in Fig. 1. Oscillograms of the responses of these hydro-

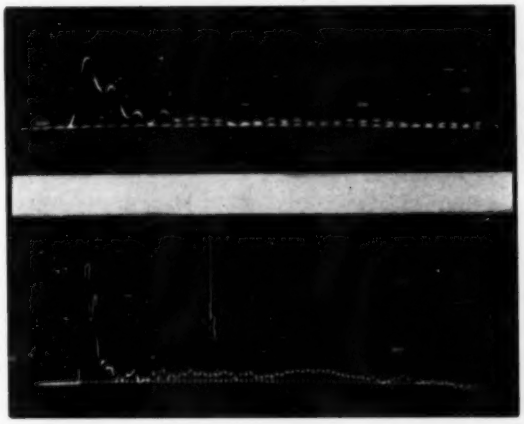


FIG. 2. Typical oscillograms of hydrophone A. The spot modulation frequency is 200 kilocycles/sec.

* Vibration Section, Sound Division.
** Mountain Lakes Laboratory, New Jersey.
[1] M. F. M. Osborne, J. App. Phys. **14**, 180–3 (April, 1943).
[2] M. F. M. Osborne and S. D. Hart, "Transmission, reflection and guiding of an exponential pulse by a steel plate in water. II. Experiment," J. Acous. Soc. Am. **18**, 170 (1946).
[3] M. F. M. Osborne and A. H. Taylor, "Non-linear propagation of underwater shock waves," Phys. Rev. **70**, 322 (1946).
[4] A. V. Bedford and G. L. Fredenhall, Proc. I.R.E. **30**, 440–447 (October, 1942).

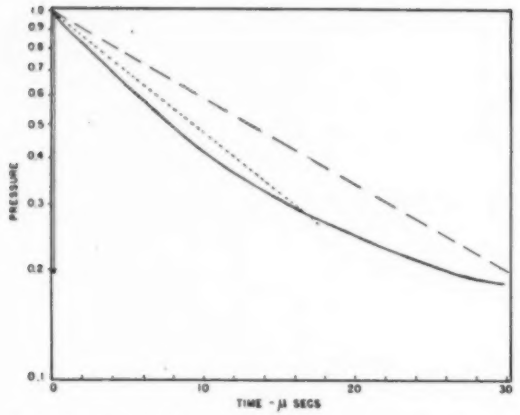


FIG. 3. The incident transient, or pressure in the explosion wave as a function of time. The ordinate units are arbitrary, being given in terms of the peak pressure as unit.

phones to the explosion wave were obtained with a television amplifier and cathode-ray oscillograph.[2] Typical oscillograms are shown in Fig. 2. The time scale is given by modulation of the spot intensity.

In order to obtain the indicial admittance of the hydrophone from the observed oscillograms, one has to know the true shape of the explosion wave. This in turn has to be obtained from oscillograms from other pressure gauges known to be flat over a frequency range much greater than the range of interest of the hydrophone under investigation. The adopted shape of the explosion wave, taken as a mean from a number of gauges,[2,3] is given in Fig. 3, full curve. The dotted and dashed straight lines are exponential approximations, $p=p_{\max}e^{-t/\tau}$.

Using the method previously described,[1] with the full curve of Fig. 3 as the incident transient, the indicial admittance (full curve) of Fig. 5 was obtained. If the straight line approximations of Fig. 3 are used as the incident transient, the analysis is considerably simplified, as follows.

If $1(t)$ is the unit function, and $H(p)$ the impedance operator of the hydrophone (here $p=d/dt$, not pressure), $A(t)$ the indicial admittance, then

$$H(p)\cdot 1(t)=A(t). \tag{1}$$

Now the impedance operator or operational equivalent for a simple exponential function is $p/(ap+b)$, so that

$$p/(ap+b)\cdot 1(t)=p_{\max}e^{-t/\tau}, \tag{2}$$

where $a=1/p_{\max}$, $b=1/p_{\max}\tau$. (This is easily shown from the differential equation of a circuit consisting of a battery, resistance, and capacitance in series.)

Hence the inverse operator, $a+b/p$, operating on the exponential explosion wave $p_{\max}e^{-t/\tau}$ converts it into the unit function

$$(a+b/p)\cdot p_{\max}e^{-t/\tau}=1(t). \tag{3}$$

Now the impedance operator $H(p)$ operating on the explosion wave gives just the observed oscillogram $O(t)$

$$H(p)\cdot p_{\max}e^{-t/\tau}=O(t).$$

Hence, using (3)

$$(a+b/p)H(p)p_{\max}e^{-t/\tau}=(a+b/p)\cdot O(t),$$

$$H(p)\cdot 1(t)=A(t)=(a+b/p)\cdot O(t).$$

One has therefore the very simple rule; to convert the response to the exponential (observed oscillogram) to the indicial admittance, one simply takes a linear combination of the observed oscillogram and its indefinite integral, i.e., the ordinates of the oscillogram and its "integral curve" (Fig. 4) are combined in constant proportion. The result of this process is shown by the dotted and dashed curves of Fig. 5, corresponding to the curves of Fig. 4. They fall below the full curve, corresponding to the overestimate of the time constant (b too small).

It will be observed, by comparing Figs. 4 and 5, that the contribution of the integrated term becomes more and more nearly a constant, so that the observed oscillogram is, after a short time, just the indicial admittance lowered by a constant amount. Essentially, the observed oscillogram is just the true indicial admittance

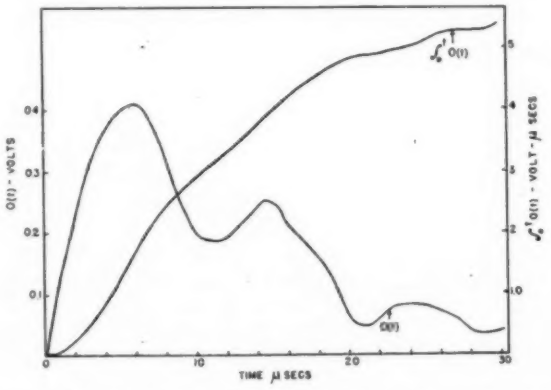


FIG. 4. The observed oscillogram for hydrophone A and its integral curve.

"wrapped around" an exponential type function. However, if the time constant of the explosion wave is *very* short compared to the characteristic periods of the structure under investigation, the observed response will be approximately the derivative of the indicial admittance.

The simple relation which exists between the indicial admittance and the response to the explosion wave suggests that the explosion wave can be used as a tool in acoustical "circuit analysis" in much the same way that a square wave generator is used in electrical circuit analysis.[2]

As a check on the consistency of the data (but not on the assumption that the explosion wave is an exponential) one can perform Fourier analyses of the oscillograms from two different hydrophones and compare the ratio of these with the ratio of the calibrations determined directly in continuous waves. This was done by graphical methods[4] and also by using a mechanical harmonic analyzer whose distinguishing characteristic was its ability to analyze a curve of arbitrary base length.[5] The result of this analysis is shown in Fig. 6 where it will be seen that the explosion wave comparison is in good agreement with the continuous wave comparison.

Since the absolute calibration of the hydrophones is known, this analysis can also give the spectrum of the explosion wave. Actually the Fourier transform of the derivative of the explosion wave is given, which is equal to the frequency time the transform of the explosion wave itself. This is shown in Fig. 6. The limiting ordinate for high frequencies is the peak pressure p_{max}. This is $\simeq 3.5\times10^6$ dynes/cm² at a distance of 18 feet (5.5 meters) from the explosion, a value in good agreement with the expected value from other more direct methods.[2,3] The dotted curve in Fig. 7 is the theoretical curve for an exponential explosion wave corresponding to the dotted curve of Fig. 5. Systematic deviations appear at the low frequency end. This end of the curve is poorly determined experimentally, since the oscillograms from which the data were obtained only lasted one or two periods of the corresponding frequency ($\sim$5 kc). The deviations of points from the simple theory are in the right direction (above the curve) since there are some theoretical reasons for thinking that the explosion wave dies off less rapidly than an exponential for large values of the time.

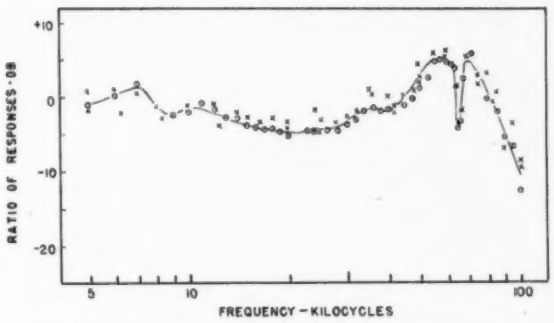


FIG. 6. Ratio of frequency calibrations of hydrophones *A* and *B*. Crosses from explosion wave data, circles from continuous wave comparison.

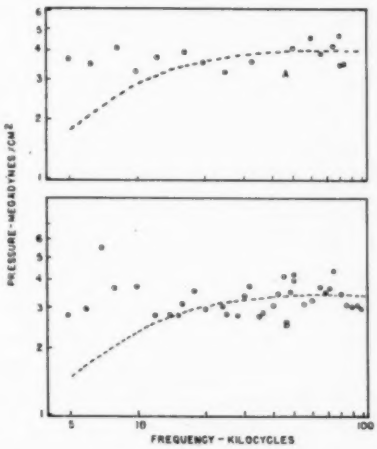


FIG. 7. "Spectrum" of explosion wave. The dotted curve is the theoretical curve for a simple exponential.

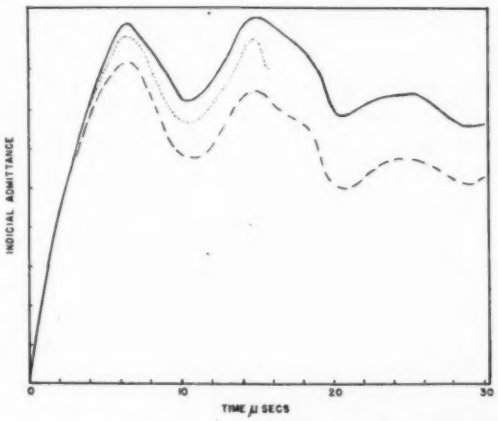


FIG. 5. The indicial admittance of hydrophone *A*, using Figs. 3 and 4. The ordinate scale is arbitrary, the units can be either open circuit volts/dyne-cm² or coulombs/dyne-cm².

[5] O. Mader, Electrotech. Zeits. **30**, 847 (September 9, 1909).

An X-Ray Method for Measuring the Thickness of Thin Crystalline Films

A. EISENSTEIN*
*Radiation Laboratory,** Massachusetts Institute of Technology, Cambridge, Massachusetts*
(Received May 31, 1946)

When x-rays are scattered from a thin crystalline surface film overlying a crystalline base material, diffraction lines are observed from both materials. Equations are developed for the general case of (1) a flat sample and (2) a cylindrical sample expressing the ratio of line intensities from the two materials as a function of film thickness. A usable range of 10^{-6} cm to 5×10^{-2} cm is indicated. Experimental confirmation is found in the range 10^{-4} cm to 6×10^{-3} cm.

INTRODUCTION

ALTHOUGH electron diffraction methods are generally used to detect and analyze thin surface films, x-ray diffraction patterns have been obtained[1] from electrodeposited films of copper as thin as 5×10^{-6} cm. In these experiments determinations of thickness were made from the time and current flow conditions of the electrodeposition process rather than from the x-ray patterns which served only to indicate the crystal structure present. Since diffraction patterns of thin surface films are composed of diffraction lines from both the underlying base material and surface layer, it seems reasonable to expect that a comparison of the relative, integrated intensities of the two sets of diffraction lines can be used as a measure of the surface film thickness. Such a comparison must take into account differences in absorption, crystal structure, and structure factor of the two materials as well as geometrical considerations. Equations are developed for use in obtaining the thickness of thin surface films on either flat or cylindrical samples. An experimental verification is found in the case of the cylindrical sample.

Since the completion of this work two abstracts have appeared describing thickness measurements of thin coatings using x-ray techniques. One of these methods,[2] based essentially on absorption, compares the intensities of certain diffraction lines of the base material observed before and after coating with the surface layer. From this measured intensity ratio, the absorption coefficient of the surface material and the geometry of the arrangement, it was possible to determine metal plating thicknesses in the range 10^{-5} cm to 10^{-2} cm. The second method,[3] essentially that which is described in this paper, compares the intensities of diffraction lines from the surface and base materials. An equation is shown for the special geometrical condition of back reflection and is verified experimentally in the thickness range of 10^{-4} cm.

FLAT SAMPLE

Figure 1 shows a collimated, homogeneous, primary beam of wave-length λ and intensity I_0 falling upon a mass of crystalline powder having a total volume V. A total power P is diffracted into the Debye-Scherrer halo forming a cone of semi-apex angle 2θ. The power which is diffracted into a circular segment of length s and is recorded by a cylindrical strip of film is given by[4]

$$P'=\frac{I_0 s e^4\lambda^3}{16\pi r m^2 c^4}N^2 V p F^2\frac{1+\cos^2 2\theta}{\sin 2\theta\sin\theta}A(\theta), \quad (1)$$

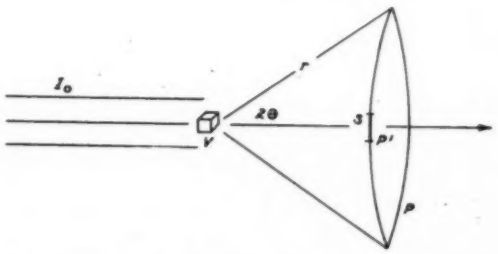


FIG. 1. Debye-Scherrer scattering from volume V.

* Now at the Physics Department, Massachusetts Institute of Technology.

** This paper is based on work done for the Office of Scientific Research and Development under contract OEMsr-262 with Massachusetts Institute of Technology.

[1] G. L. Clark, G. Pish, and L. E. Weeg, J. App. Phys. **15**, 193 (1944).

[2] L. S. Birks and H. Friedman, Bull. Am. Phys. Soc. **20**, No. 4 (1945).

[3] R. B. Gray, Bull. Am. Phys. Soc. **20**, No. 5 (1945).

[4] *International Tables for the Determination of Crystal Structures* (Gebrüder Borntraeger, Berlin, 1935), p. 561.

where

N = number of unit cells per unit volume or $1/Vc$ where Vc is the unit cell volume,
p = multiplicity,
F = crystal structure factor,
r = camera radius, and
$A(\theta)$ = correction for absorption in the sample.
e, m, and c have their usual significance.

The power dP_i' diffracted from the volume element dV_i is,

$$dP_i' = KN_i^2 p_i F_i^2 (\text{L.P.})_i A(\theta)_i dV_i, \quad (2)$$

where

$$K = (I_0 s e^4 \lambda^3) \div (16\pi r m^2 c^4), \quad (3)$$

and includes those terms which are independent of the particular volume element chosen. (L.P.) designates the usual Lorentz-Polarization factor.

The geometrical arrangement for the general case of diffraction from a double layer flat sample is shown in Fig. 2. A primary beam of intensity I_0 and cross section A_0 falls on the sample at an incident angle α irradiating a projected area A. This gives rise to a diffracted power dP_1' from the volume element dV_1 of layer 1 and to dP_2' from a similar volume element of layer 2. Radiation scattered from dV_1 must traverse a total path length in layer 1 of xy_1 where

$$xy_1 = x\left(\frac{1}{\sin\alpha} + \frac{1}{\sin(2\theta_1 - \alpha)}\right). \quad (4)$$

Introducing the absorption correction and

$$dV_1 = A\,dx, \quad (5)$$

$$dP_1' = KAN_1^2 p_1 F_1^2 (\text{L.P.})_1 \exp\{-\mu_1 x y_1\} dx, \quad (6)$$

where μ_1 is the linear absorption coefficient and F_1 the crystal structure factor for the material of this layer. By integration between the limits 0 and l, the power scattered from layer 1 of thickness l,

$$P_1' = KAN_1^2 p_1 F_1^2 (\text{L.P.})_1 \frac{(1 - \exp\{-\mu_1 l y_1\})}{\mu_1 y_1}. \quad (7)$$

Consider the scattering from the volume element dV_2 of layer 2 which is located at a depth greater than l below the surface. The total path traversed by the diffracted rays is

$$l y_2 + x y_2 \quad (8)$$

in layers 1 and 2, respectively. By substitution

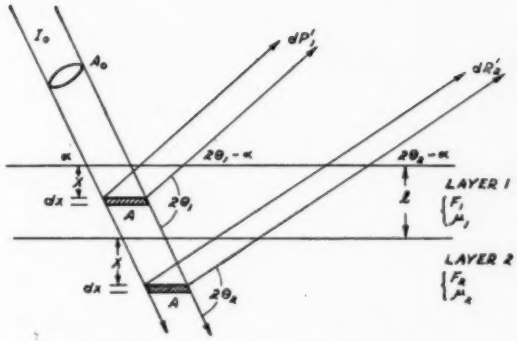


FIG. 2. Diffraction from double layer flat sample.

in (2) the diffracted power is,

$$dP_2' = KAN_2^2 p_2 F_2^2 (\text{L.P.})_2 \exp\{-\mu_1 l y_2\} \exp\{-\mu_2 x y_2\} dx. \quad (9)$$

If the absorption coefficient μ_2 is made large, either by a suitable choice of backing material or primary beam wave-length, layer 2 may be considered as having an infinite thickness for purposes of integration. Hence, integration between the limits 0 and ∞ yields,

$$P_2' = KAN_2^2 p_2 F_2^2 (\text{L.P.})_2 \frac{\exp\{-\mu_1 l y_2\}}{\mu_2 y_2}. \quad (10)$$

The ratio of integrated intensities from the materials of layers 1 and 2 is (7)/(10) or

$$\frac{P_1'}{P_2'} = \frac{N_1^2 p_1 F_1^2 \mu_2 y_2 (\text{L.P.})_1}{N_2^2 p_2 F_2^2 \mu_1 y_1 (\text{L.P.})_2} \frac{1 - \exp\{-\mu_1 l y_1\}}{\exp\{-\mu_1 l y_2\}}. \quad (11)$$

When it is convenient to select for comparison, diffraction lines from the two materials falling at approximately the same angular position, the ratio (11) may be simplified. If $\theta_1 = \theta_2$, then $y_1 = y_2$ and $(\text{L.P.})_1 = (\text{L.P.})_2$ and

$$l = \frac{1}{\mu_1 y} \ln\left\{\frac{P_1' N_2^2 p_2 F_2^2 \mu_1}{P_2' N_1^2 p_1 F_1^2 \mu_2} + 1\right\}. \quad (12)$$

The thickness of the surface material may thus be determined from measurements of the integrated intensities of selected diffraction lines from the two layers. A knowledge is also required of the unit cell sizes and the linear absorption coefficients of the two materials as well as the crystal structure factors and multiplicities for the particular *hkl* reflections.

If the materials of layers 1 and 2 have approximately the same N, p, F, and μ values, the

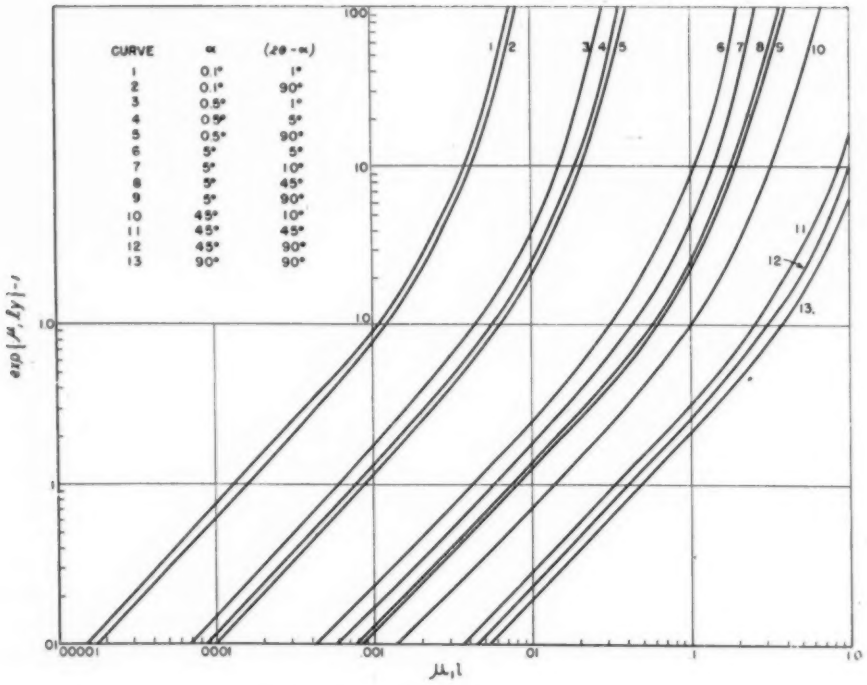


FIG. 3. (exp $\mu_1 ly-1$) *vs.* $\mu_1 l$, see text.

ratio of diffracted powers P_1'/P_2' is simply (exp $\{\mu_1 ly\}-1$). A plot of this exponential function *vs.* $\mu_1 l$ is shown in Fig. 3 for various values of the incidence angle α and the reflection angle $(2\theta-\alpha)$. Because of the interchangeability of α and $(2\theta-\alpha)$ in this function, these curves cover the experimental range from glancing incident angle, 0.1°, to normal incidence angle, 90°, as used in the back reflection x-ray method. In general, the variation of the exponential function with the reflection angle $(2\theta-\alpha)$ is small for a given incidence angle α, hence the diffraction lines selected for comparison need only fall in the same general angular range. Assuming the ratio P_1'/P_2' can be measured between 0.01 and 100, the experimental limits of thickness measurements by this method are seen to depend critically upon the α and $(2\theta-\alpha)$ values selected.

As an example of the application of this method, consider the simple case of a surface layer which has the same crystal structure as that of the underlying material but different lattice constant, atomic structure factor f, and absorption coefficient, e.g., silver over copper. Selecting the 400 line of Ag and the 222 line of Cu which fall at about the same angular position, $2\theta=95°$, Table I shows values to be substituted in (12). Atomic structure values as well as linear absorption coefficients are given for the radiation wave-length of Cu$K\alpha$, 1.54A. Assuming the ratio P_1'/P_2' measurable between 0.01 and 100, a substitution of these values into (12) leads to a measurable thickness range for glancing incidence, $\alpha=0.5°$, of 10^{-7} cm to 2×10^{-5} cm and for back reflection, $\alpha=90°$, of 10^{-6} cm to 2×10^{-4} cm. The lower limit of thickness measurement could be extended, in the case of glancing incidence, by the selection of lines falling at a smaller value of 2θ, e.g., 111 of Cu and 200 of Ag. Irregularities in the surface will tend, however, to increase the practical lower limits of measurement.

CYLINDRICAL SAMPLE

Figure 4 shows a cylindrical sample of radius R and length L intercepting a primary beam of collimated radiation and scattering a diffracted

TABLE I.

	N	p	f	$F=4f$	μ
Ag (400)	$1/(4.01)^3$	6	24.8	99.2	2280
Cu (222)	$1/(3.61)^3$	8	13.7	54.8	455

beam at an angle 2θ with the incident direction. When the absorption coefficient of either layer is large, only the surface areas AB and $A'B'$ contribute to the diffracted beam. The total diffracted power P'_{Tk}, falling in the segment s of Fig. 1, can be represented as the summation of scattering dP'_k from flat surface elements dA integrated over the effective scattering surface AB. Then,

$$P'_{Tk}=\int_A^B dP'_k, \tag{13}$$

where dP'_k is obtained by substituting dA for A in (7) or (10). Since the surface element

$$dA=LRd\alpha, \tag{14}$$

substitution into (13) yields,

$$P'_{T1}=\frac{KLRN_1^2p_1F_1^2(\text{L.P.})_1}{\mu_1}\times\int_0^{2\theta_1}\frac{1-\exp\{-\mu_1 l y_1\}}{y_1}d\alpha, \tag{15}$$

and

$$P'_{T2}=\frac{KLRN_2^2p_2F_2^2(\text{L.P.})_2}{\mu_2}\times\int_0^{2\theta_2}\frac{\exp\{-\mu_1 l y_2\}}{y_2}d\alpha; \tag{16}$$

and the ratio (15)/(16),

$$\frac{P'_{T1}}{P'_{T2}}=\frac{N_1^2p_1F_1^2\mu_2(\text{L.P.})_1}{N_2^2p_2F_2^2\mu_1(\text{L.P.})_2}\times\frac{\int_0^{2\theta_1}\frac{1}{y_1}(1-\exp\{-\mu_1 l y_1\})d\alpha}{\int_0^{2\theta_2}\frac{1}{y_2}\exp\{-\mu_1 l y_2\}d\alpha}. \tag{17}$$

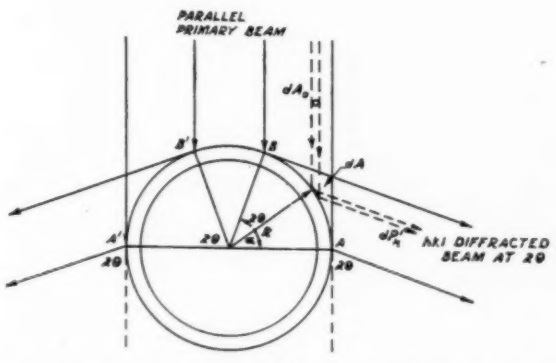


FIG. 4. Diffraction from double layer cylindrical sample.

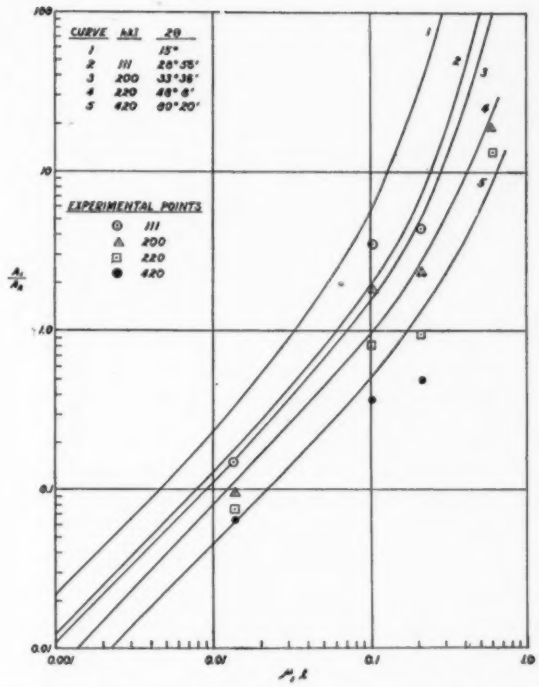


FIG. 5. Ratio of integrals of (15) and (16), A_1/A_2, respectively, *vs.* $\mu_1 l$.

TABLE II.

hkl	111	200	220	420
2θ	28°51′	33′36′	48°8′	80°20′

The integrals of (15) and (16) have been solved only by graphical integration and these values are designated as A_1 and A_2, respectively.

Selecting again for comparison, lines in the same angular range $\theta_1=\theta_2$, then $y_1=y_2$ and $(\text{L.P.})_1=(\text{L.P.})_2$, hence

$$\frac{A_1}{A_2}=\frac{P'_{T1}N_2^2p_2F_2^2\mu_1}{P'_{T2}N_1^2p_1F_1^2\mu_2}. \tag{18}$$

The ratio A_1/A_2 was evaluated as a function of $\mu_1 l$ for specific values of 2θ. The exact values selected were 15° and those which correspond to the angular positions of diffraction lines from a (BaSr)O equal molar solid solution for Cu $K\alpha$ radiation, see Table II. The ratio $A1/A2$ is shown as a function of $\mu_1 l$ in Fig. 5. This set of curves for the case of a cylindrical sample, are seen to correspond roughly to the set in Fig. 2 for a flat sample set at an incident, glancing angle of 5°.

In practice, the ratio A_1/A_2 is obtained by means of (18) from the experimentally measured, integrated intensity ratio P'_{T1}/P'_{T2} and the thickness l determined from Fig. 5. As an example, consider again the case of an Ag layer covering a Cu base structure and compare the diffracted 400 and 222 lines, respectively. Assuming the ratio P'_{T1}/P'_{T2} can be measured between 0.01 and 100, substitution into (18) indicates a measurable thickness range of 3×10^{-6} cm to 10^{-3} cm. The selection of diffraction lines occurring at smaller 2θ positions would decrease the limits of this range, as indicated in Fig. 5.

An inspection of (7) and (10) shows that the diffracted power dP'_k per unit area dA of periphery is symmetrical with respect to variations in α, about $\alpha=\theta$. Hence, there is an asymmetrical distribution of intensity, power per unit area of diffracted beam, across the beam as it is intercepted by the film. This effect is of interest in assigning an appropriate correction for the angular displacement of diffraction lines and has been discussed[5,6] in this connection.

EXPERIMENTAL

A series of cylindrical thickness standards were prepared to establish experimentally the validity of Eq. (18). The samples used were in the form of indirectly heated oxide cathodes. Nickel cylinders 3 mm in diameter were sprayed with a coating of $BaCO_3$, sufficiently thick to prevent scattering from the underlying base metal. Over these were sprayed thin, uniform coatings of $SrCO_3$ of differing weights. By heating in vacuum at an uncorrected pyrometer temperature of 800°C, the carbonates were converted to the oxides BaO and SrO. At this temperature no interdiffusion of the two layers to form a solid solution could be detected in the x-ray patterns of the wax protected cathodes.[7] These patterns were obtained using filtered Cu $K\alpha$ radiation, a 3×0.5-mm slit collimator and a 4.70-cm radius camera. Microphotometer tracings were made of the lines diffracted from the 111, 200, 220, and 420 planes of SrO and BaO. As both materials have the same crystal structure and differed by only 15 percent in lattice constant, the hkl lines of SrO were compared directly with the corresponding lines of BaO which fell in approximately the same angular range. The ratio P'_{T1}/P'_{T2} was simply the ratio of maximum line intensities I_1/I_2 since the shape of all diffraction lines appeared to be the same. From the spray conditions, the apparent density of the $BaCO_3$ and $SrCO_3$ layers was known to be 1.2 ± 0.2 g/cm³ and 0.8 ± 0.2 g/cm³, respectively. SrO coating thicknesses were computed from the apparent $SrCO_3$ density, the weight of each coating, and the coated area. These were found to be 1.3×10^{-4} cm, 1.0×10^{-3} cm, $2.0\times.0^{-3}$ cm, and 6.3×10^{-3} cm. No correction was made for the small reduction in thickness and increase in density produced in both layers during the conversion process. The experimental points, obtained from (18) with $p_1=p_2$ and appropriate N, F, and μ values, for the oxides using the apparent density values, fall on the $A1/A2$ curves as indicated in Fig. 5.

This method of measuring film thickness has been used to determine, (1) the thickness of electrodeposited layers, (2) the extent of surface oxidation after the exposure of the sample to an oxidizing atmosphere, (3) the thickness of sprayed oxide coatings, and (4) the thickness of surface layers produced by a chemical reaction.

The author is indebted to Professor B. E. Warren for his very helpful criticism of this manuscript as well as to Miss Eleanor Uhl for the evaluation of the function presented in Fig. 5.

[5] A. Taylor and H. Sinclair, Proc. Phys. Soc. **57**, 108 (1945).
[6] B. E. Warren, J. App. Phys. **16**, 614 (1945).
[7] A. Eisenstein, J. App. Phys. **17**, 434 (1946).

A New Bridge Photo-Cell Employing a Photo-Conductive Effect in Silicon. Some Properties of High Purity Silicon*[1]

G. K. Teal, J. R. Fisher, and A. W. Treptow
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

(Received July 2, 1946)

A pure photo-conductive effect was found in pyrolytically deposited and vaporized silicon films. An apparatus is described for making bridge type photo-cells by reaction of silicon tetrachloride and hydrogen gases at ceramic or quartz surfaces at high temperatures. The maximum photosensitivity occurs at 8400–8600A with considerable response in the visible region of the spectrum. The sensitivity of the cell appears about equivalent to that of the selenium bridge and its stability and speed of response are far better. For pyrolytic films on porcelain there are three distinct regions in the conductivity as a function of temperature. At low temperatures the electronic conductivity is given by the expression $\sigma = Af(T) \exp -(E/2kT)$. At temperatures between 227°C and a higher temperature of 400–500°C $\sigma = A \exp -(E/2kT)$, where E lies between 0.3 and 0.8 ev; and at high temperatures $\sigma = A \exp -(E/2kT)$, where $E = 1.12$ ev. The value 1.12 ev represents the separation of the conducting and non-conducting bands in silicon. The long wave limit of the optical absorption of silicon was found to lie at approximately 10,500A (1.18 ev). The data lead to the conclusion that the same electron bands are concerned in the photoelectric, optical, and thermal processes and that the low values of specific conductances found (1.8×10^{-5} ohm^{-1} cm^{-1}) are caused by the high purity of the silicon rather than by its polycrystalline structure.

I. INTRODUCTION

IT has long been known that the electrical conductivity of certain solid insulating or semi-conducting materials can be temporarily increased by illumination with light of suitable wave-length.[2] Several such materials have been employed in resistance bridge type of photo-cells and have been commercially available. The conductivity imparted by light is electronic in character and results from a photoelectric effect occurring within the individual crystals. Two types of crystals have been recognized in studies of the photo-conductivity of various solids. Some materials exhibit photo-conductivity when chemically pure and in the form of perfect or nearly perfect single crystals. These are known as idiochromatic crystals; typical examples are diamond, zincblende, sulphur, the red variety of selenium, stibinite, and cinnabar. Other materials show no photo-conductivity in the pure state and acquire it only through incorporation of impurities into the regular crystal lattice, by irradiation by x-rays or ultraviolet light, by application of high electric fields, or by thermal treatment to displace constituents from their normal lattice positions or partially to decompose the material so as to result in a stoichiometric unbalance of the normal lattice constituents. Such materials are known as allochromatic crystals and include the alkali, silver, and thallium halides. There are, in addition to the particular materials mentioned, a great number of other materials that have been investigated in polycrystalline form and found to be photo-conducting but which have not been studied sufficiently to permit their classification.

The process of conduction has been shown to be the sum of a primary effect and secondary effects. The primary process is instantaneous whereas the secondary process is the result of the passage of current and may take many minutes before reaching an equilibrium state. It is of considerable practical import for some purposes, therefore, that the material employed in a resistance bridge photo-cell shall have a highly sensitive primary current accompanied by little or no secondary current.

The best known commercial bridge photo-cells are perhaps the selenium and thalofide cells. The photo-conductive effect in the selenium cell is accompanied by large secondary effects. The

* This paper is based in whole or in part on work done for the Office of Scientific Research and Development under Contract OEMsr-1231 with the Western Electric Company.

[1] Presented at meeting of the Am. Phys. Soc. April 26, 1946. Phys. Rev. **69**, 686A (1946). See also NDRC Sec. 16.4–37 Report, Jan. 4, 1945, G. K. Teal.

[2] For summary see A. L. Hughes and L. A. DuBridge, *Photoelectric Phenomena* (McGraw-Hill Book Company, Inc., New York, 1932), Chap. VIII; F. C. Nix, Rev. Mod. Phys. **4**, 723 (1932); A. L. Hughes, Rev. Mod. Phys. **8**, 294 (1936).

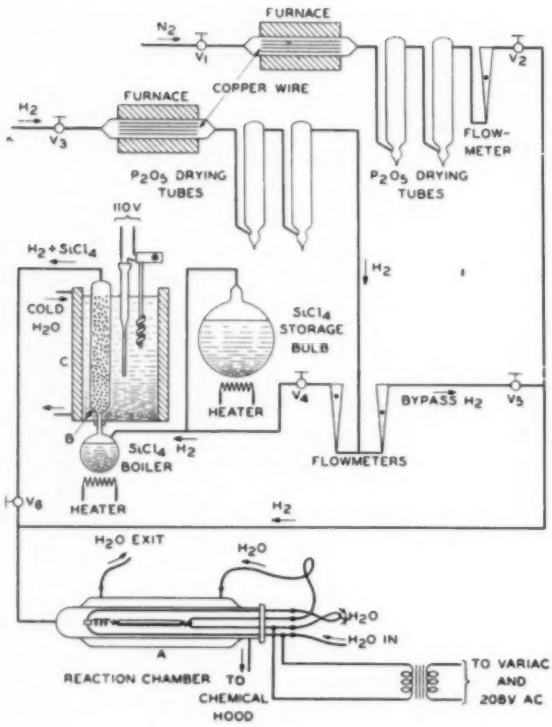


FIG. 1.

secondary effects in thalofide, however, do not appear to be so serious.

During the war several investigators observed photo-effects in silicon. In the present paper we shall describe a pure photo-conductivity found in silicon[3] and a method of preparing bridge photo-cells employing the effect. Experimental evidence will be presented tending to prove that silicon is an idiochromatic crystal with minor secondary effects. Several attractive features of the cells will be described.

II. METHODS OF PREPARING CELLS

Silicon cells may be prepared by use of pyrolytically deposited or vaporized silicon. The cells described below were made by pyrolytic deposition of silicon on ceramic and quartz bases. For this purpose the reduction of silicon tetrachloride vapors by hydrogen gas at a hot surface[4] is satisfactory. It seems likely that this reaction is a complicated one involving the formation of intermediate products of long chain silicon chlorides.[5] For the purposes of simplicity, however, we shall write it as:

$$SiCl_4+2H_2\rightarrow Si+4HCl.$$

One method of preparing cells has been to deposit the silicon films on ceramic tubes suspended on internal tantalum wire heaters in flowing silicon tetrachloride and hydrogen gases. A schematic diagram of the experimental apparatus is shown in Fig. 1. It consists of a gas flow system for providing controlled amounts of hydrogen, nitrogen, and silicon tetrachloride to the water cooled reaction chamber *A* in which the ceramic tube is suspended by means of the internal wire heater, the ends of which are connected to water cooled copper electrical leads. As can be seen in the diagram the purified hydrogen may be introduced into the reaction chamber through a by-pass and also through a silicon tetrachloride supply branch. Each branch is provided with a floating ball flow-meter. The amount of silicon tetrachloride introduced into the reaction chamber is controlled by the hydrogen flow through the silicon tetrachloride boiler and also the temperature of the reflux condenser *B*, which is surrounded by the hollow-walled metal thermostat *C*. Cells have also been made by heating the ceramic or quartz base in a furnace tube connected to the gas supply system in place of the reaction chamber *A*. The temperature of the ceramic base is determined by means of an optical pyrometer when the internal wire heater is used and by means of a thermocouple when the furnace is employed.

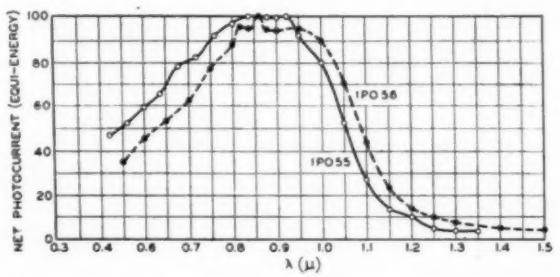


FIG. 2. Spectral sensitivity of silicon bridge photo-cells.

[3] Since preparing the present paper two reports on other photoeffects in silicon have come to our attention. P. H. Miller, Jr. and M. H. Greenblatt, U. of Pennsylvania OSRD Report No. 412, March 20, 1945; and F. C. Brown, Phys. Rev. **69**, 686A (1946).

[4] R. Hölbling, Zeits. f. angewandte Chemie **40**, 655 (1927); J. N. Pring and W. Fielding, J. Chem. Soc. **95**, 1497 (1909).

[5] R. Schwartz and H. Meckback, Zeits. f. anorg. u. allgem. Chemie **232**, 241 (1937); R. Schwartz and R. Thiel, *ibid.* **235**, 247 (1938).

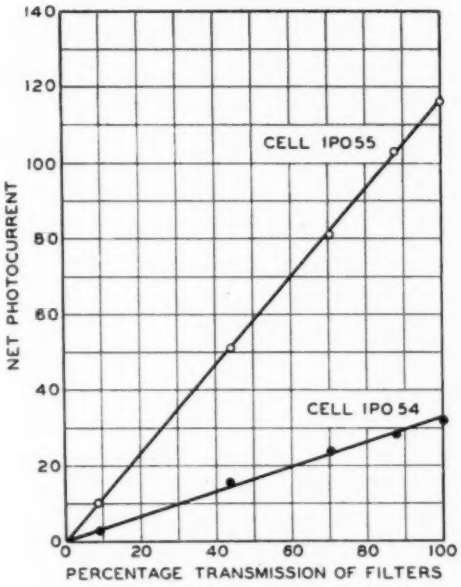


FIG. 3. Dependence of the photo-response to λ9200 on intensity of illumination.

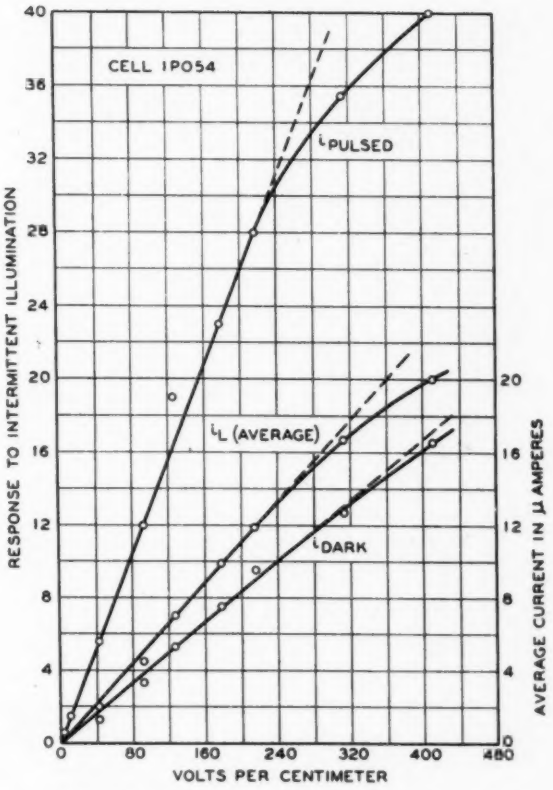


FIG. 4. Linearity of response with applied voltage gradient.

One type of cell consists of a thin film of silicon on the outside of a ceramic rod about 95 mils in diameter and about ¾ inch long. Another is a thin film of silicon on the flat surface of quartz or ceramic of several square inches area. The electrodes may be silver paste fired on at 600–800°C or a conducting silver paint applied and air dried at low temperatures.

Using the means just described silicon deposits readily on ceramic surfaces heated to about 1000°C or higher. The deposits made at the lower temperatures tend to be grainier than those made at the higher temperatures. In many cases the deposit made at the lower temperatures appears to the eye as a brown fuzz. Under the microscope this fuzz is seen to be long hairlike single crystals. At temperatures of about 1100°C and above the silicon deposits as a smooth, hard, gray film adhering very firmly to the ceramic base. A wide range of film thicknesses have been made. In the case of a typical film 5 microns thick a deposition time of 1.5 minutes suffices. Under a given set of gas flow conditions the rate of deposition at 1250°C has been observed to be about 5 times greater than the rate at 1000°C. The film thickness is a linear function of the deposition time. Rate of deposition is approximately proportional to the silicon tetrachloride concentration up to a certain concentration above which the rate decreases.

III. CHARACTERISTICS OF THE PHOTO-CONDUCTIVE EFFECT

In Fig. 2 are shown spectral sensitivity curves of two typical cells. The maximum for each cell is set equal to 100 on an arbitrary current scale since the absolute magnitude of the photo-effect in individual cells has not been determined. However, for each separate cell the responses to light of different wave-length are on an equal energy basis. The response curves are seen to correspond closely, the maximum response in all cases lying between 8400 to 8600A. The long wave-length limits were not determined with great accuracy but appear to lie between 12,000 to 15,000A. The photo-sensitivity in all cases extends well into the visible region and decreases steadily as the wave-length decreases.

In order to be certain that the broadness of the spectral maxima was not due to saturation effects, the photo-conductive responses of several

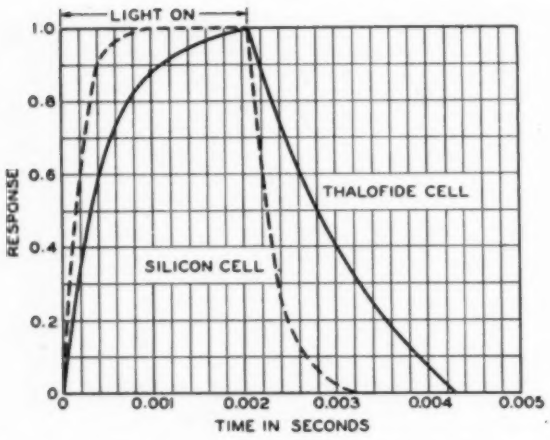


FIG. 5. Dynamic responses of silicon and thalofide cells to intermittent illumination.

cells were determined as a function of the light intensity for $\lambda = 9200$A under the same conditions used in obtaining the spectral sensitivity curves. With tungsten light having a color temperature of 2848°K the maximum photo-current (uncorrected for equal energy) was obtained at about $\lambda = 9200$A. Levy line screens were used as filters and their transmissions were determined in the same optical setup that was used in making the measurements on the silicon cells. The results are shown, in Fig. 3, for two cells. In each case the photo-current is a linear function of the light intensity. The great breadth of the sensitivity curves near the maximum is therefore not attributable to saturation effects depressing the curve at the maximum.

In Fig. 4 there are shown plots of the response of one of the rod shaped cells to intermittent illumination as observed on a cathode-ray oscillograph. The light was interrupted by a light chopper at the rate of about 240 cycles per second. The cell was coupled to the oscillograph through a 10,000-ohm resistor. In addition the average dark and light currents were measured by means of a meter. The curves show that the a.c. response is linear with the field applied up to about 200 volts per cm. Measurements on other cells likewise indicated linearity of response with increasing field applied to the film.

The sensitivities of several rod shaped cells to uniform illumination are given in Table I. It was found that such a condition could be approximated by placing the rod cells at the center of an equilateral triangle, at whose corners are situated three frosted bulb lamps. 40-watt bulbs were used on a regulated voltage supply. The illumination was determined by substituting a Weston photometer for the silicon cells and averaging for the three light sources. The illumination was thus found to be 127 footcandles. The potential across the cells was 48 volts in all cases. The sensitivity observed when the area illuminated is small in comparison to the diameter and length of the rod is somewhat lower than the above values and shows some variation with position of the area on the rod. The lower values, however, appear not to be associated with saturation effects due to large flux densities since observations show that the response varies linearly with the total light intensity for small spots and for the highest flux densities used. The highest sensitivities have been obtained for cells made in the glass reaction chamber of Fig. 1. The sensitivity of the silicon cells appears considerably lower than that obtainable in the thalofide cell but is considerably greater than that of a Luxtron selenium cell that we have had on hand for several years.

Some idea of the relative speeds of response of the silicon and thalofide cells can be gained from the approximate curves shown in Fig. 5. The curves were obtained on a cathode-ray oscillograph while illuminating the cells with an intermittent light signal of about 240 cycles per second. The scales are arbitrary. Though the thalofide cell shows a larger response than the silicon cell the speed of response is seen to be definitely faster in the case of silicon. The time of attainment of an equilibrium state on starting or stopping the illumination of a silicon cell is less than 0.001 second under the conditions used. In selenium the equilibrium state may not be reached for many minutes. Measurements of the response of some of the silicon cells to an

TABLE I. Sensitivity of silicon cells with uniform illumination.

Cell No.	Microamperes per lumen per 200 volts per cm	Percent change in dark current per lumen	i_l/i_d per lumen	i_l/i_d for 127 ft. candles
1P055	55.8	1720	18.2	2.007
$20518F_5$	74.0	1890	19.9	2.054
$20518F_6$	34.0	1040	11.4	2.193

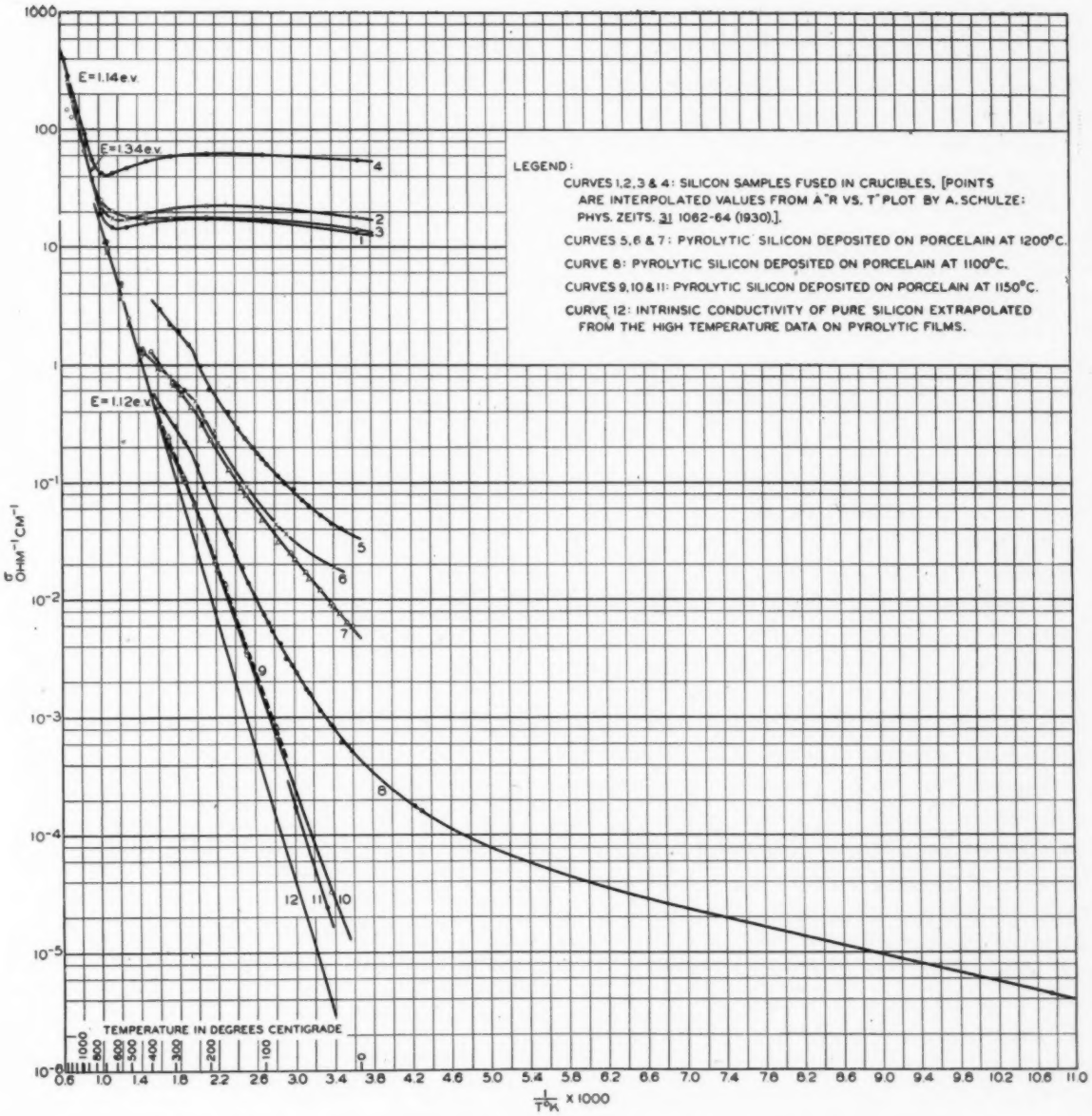

LEGEND:
CURVES 1,2,3 & 4: SILICON SAMPLES FUSED IN CRUCIBLES, [POINTS ARE INTERPOLATED VALUES FROM A"R VS. T° PLOT BY A. SCHULZE: PHYS. ZEITS. 31 1062-64 (1930).].
CURVES 5,6 & 7: PYROLYTIC SILICON DEPOSITED ON PORCELAIN AT 1200°C.
CURVE 8: PYROLYTIC SILICON DEPOSITED ON PORCELAIN AT 1100°C.
CURVES 9,10 & 11: PYROLYTIC SILICON DEPOSITED ON PORCELAIN AT 1150°C.
CURVE 12: INTRINSIC CONDUCTIVITY OF PURE SILICON EXTRAPOLATED FROM THE HIGH TEMPERATURE DATA ON PYROLYTIC FILMS.
E=1.14e.v.
E=1.34e.v.
E=1.12e.v.
σ OHM⁻¹CM⁻¹
TEMPERATURE IN DEGREES CENTIGRADE
1/T°K x 1000

FIG. 6. Plot of the logarithm of the specific conductivity of silicon films *versus* $1/T°K$.

intermittent light signal showed that the response at 720 cycles is 53 to 68 percent of the response at 90 cycles.[6] For one type of thalofide cell the 720-cycle response drops to about 14 percent of the 90-cycle response.[6]

Observations[6] of the effect of background radiation on the silicon cell indicate that it is far superior to the thalofide cell in this respect. Thus for one silicon cell studied the response to intermittent illumination was down to 72 percent for 30 footcandles background and to 25 percent for 500 footcandles. For the thalofide cell the responsivity is down to 0.1 percent for 500-footcandles background illumination.

Silicon cells heated to a red heat in an oxygen-gas flame showed no decrease in sensitivity on cooling to room temperature. No other cell that we know of could stand such treatment. Silicon cells have been left in strong sunlight for one hour without noticeably deteriorating the response.

[6] We are indebted to Drs. W. L. Hole and L. R. Posey, Jr., of the University of Michigan for these observations.

IV. OTHER ELECTRICAL PROPERTIES OF HIGH PURITY SILICON

Relatively few years ago the sign of the temperature coefficient of resistance of silicon was in dispute. One view was that the negative temperature coefficient often observed must be caused by thin films or other effects occurring at the intergranular boundaries of the polycrystalline material. From work done in recent years, however, it is generally accepted that silicon is a semi-conductor and that in the pure state it has a negative temperature coefficient of resistance. That a dispute arose is not surprising in view of the difficulty of preparing very pure silicon in compact form, the scarcity of experimental data on the quite enormous effects that even a few thousandths of a percent of an impurity can have on the electrical properties of a semiconductor, and also the fact that the modern theories of the solid state had not then been developed.

In Fig. 6 are shown experimental data on the effect of temperature on the electrical conductivity of a number of samples of silicon. The logarithm of the specific conductivity is plotted as a function of the reciprocal of the absolute temperature. Curves 1, 2, 3, and 4 are plotted from data on fused silicon obtained by Schulze.[7] They are of the same general shape and give about the same magnitudes of specific conductivity obtained in more recent samples of fused silicon. The interpretation of these curves is evidently the same as that given by Seitz[8] and Lark-Horowitz and Johnson[9] in explaining similar but more recent data on silicon and germanium. Thus the conductivity at temperatures lower than about 500°C is primarily caused by electron transitions between impurity centers and adjacent bands while the conductivity at temperatures greater than about 600°C is primarily caused by electron transitions from the lower filled band to the higher unfilled band, and the maximum and minimum arise due to the inability of the impurities and the filled band to supply conductors fast enough to make up for the decrease in conductivity due to lattice scattering. At high temperatures, as may be seen in the figure, the points fall on a straight line in the log σ *versus* $1/T$ plot. The slope of the line gives the energy separation E of the filled and unfilled bands in the solid according to the following equation:

$$\sigma = A \exp\left[-(E/2kT)\right],$$

where k is Boltzmann's constant, T is the absolute temperature, and A is a factor that varies rather slowly with temperature. Curves 1 and 2 coincide very well at high temperatures and lead to a value of 1.34 electron volts for the band separation. Curve 4 gives a value of 1.14 ev, which is in better agreement with Seitz's value of 1.12 electron volts. Curve 3 agrees somewhat better with the higher value from 1 and 2.

The silicon films prepared in the present work show entirely different orders of conductivity from fused silicon samples. This is shown in curves 5, 6, 7, 8, 9, 10, and 11 taken on various silicon films deposited pyrolytically on porcelain. The curves show three distinctly different regions: A straight line, high temperature region, a straight line intermediate temperature region, and a curved, low temperature region which appears to approach a constant slope at -100 to -180°C. The curvature is the reverse of that observed for fused silicon. There are thus two points of change one at about 227°C and another at 400° or higher, the exact temperature, in the latter case, depending on the magnitude of the conductivity. The points of change might be correlated with those previously indicated in the work of Hölbling[10] (260°, 410°, 550°) and the discontinuous and reversible changes observed at 210° and 430°C in fused silicon by Köenigsberger and Schilling.[11] The general shape of the curves is similar to an approximate result on a sample of silicon on carbon prepared by Hölbling but the conductivities in the present work are very much lower. Also the log σ *vs.* $1/T$ plot given here makes clear that there are three regions involved and indicates why some variation may occur in the temperature at which the upper inflection point is observed. Köenigsberger

[7] A. Schulze, Physik. Zeits. **31**, 1062–4 (1930).
[8] F. Seitz, OSRD report. PB2619, 20 pp. (1942). Office of the Publication Board of the Dept. of Commerce.
[9] K. Lark-Horowitz and V. A. Johnston, Phys. Rev. **69**, 258A (1946).
[10] R. Hölbling, reference 4.
[11] J. Köenigsberger and K. Schilling, Ann. d. Physik **32**, 179 (1910).

and Schilling ascribed the changes they noted to polymorphic transitions. However, such polymorphic changes have not been reported in the x-ray literature and Schulze[7] was unable to find evidence for such changes by investigation of a number of fused silicon samples up to 1000°C with an optical dilatometer. It therefore seems probable that the changes are to be correlated with changes in the conduction mechanism.

Below 227°C the curves have a changing slope, and A in the above equation must be replaced by $Af(T)$.[12] At the lowest temperature studied the curves appear to approach a constant slope equivalent to an activation energy of about 0.08 ev. It was found also that the data of curves 5, 6, 8, 9, and 10 fit very well an empirical equation of the type $\sigma=\sigma_0{}^{\alpha T}$ such that a log σ *versus* T plot gives points approximating closely a straight line. In the range of very low temperatures, 100–200°K, which have received only minor study, some deviation from the straight line may occur. A somewhat more reasonable relationship might be found in the fact that log σ *versus* log T plots of the data of curves 6, 7, 8, 9, and 10 give good fits to straight lines between about 280–500°K. Some significance might attach to the fact that these lines appear to converge on a point at about 665°K. In the case of curve 5, however, this does not hold and the departure from the straight line occurs at about 380°K, such that the conductivity appears higher than would occur due to maintenance of the straight line relationship. We have no adequate interpretation of these results at the present time.

The values of activation energy for the intermediate region 227° to 500°C vary, individual values of 0.31, 0.45, 0.49, and 0.79 electron volts being obtained.

In the high temperature region individual values of E obtained are 1.13 ev, 1.106 ev, and 1.12 ev. The straight line 12 is the best fit that can be drawn for all the high temperature measurements. It represents a band separation in silicon of 1.12 electron volts, in good agreement with Seitz and with curve 4 from Schulze's work. The values of conductance are in better agreement with the data of Seitz than that of Schulze. The high temperature data lead to a calculated value of 3.9×10^{-6} ohm^{-1} cm^{-1} for the specific conductivity of pure silicon at 25°C. The lowest specific conductivity observed in the present work was 1.8×10^{-5} ohm^{-1} cm^{-1}. For this sample the activation energy of 0.97 electron volt was the highest value observed at room temperature.

V. SUMMARY

The best photo-conducting cells are the ones showing the lowest specific conductivity, corresponding to very pure silicon. The long wave limit of the photo-conductive effect was not measured with great accuracy but appears to lie between 12,000 to 15,000A. The thermal activation energy of 1.12 electron volts corresponds to a $\lambda_0=11{,}000$A. This is considered strong evidence for the same electron bands being concerned in both the photoelectric and thermal processes of increasing the conductance. The maximum photoelectric sensitivity occurs at 8400–8600A, equivalent to 1.45 electron volts energy. It seems certain that the process is not simply a transition between an impurity level and adjacent band since this would result in a much lower value of λ_0 (in ev) than was observed.

Observations on evaporated films of silicon indicated a low value of optical absorption in the infra-red region. This is in agreement with expectations.[13] Appreciable absorption in the shorter wave-lengths starts at about 10,500A, increasing rapidly with decreasing wave-length until a relatively high value is attained in the neighborhood of 8750A. The high absorption continues into the visible part of the spectrum. The approximate value 10,500A for the long wave limit of the absorption, corresponds to a separation of the conducting and non-conducting bands of silicon equivalent to 1.18 electron volts. Such evaporated films also show photo-conductance. The maximum photoelectric sensitivity of pyrolytic films is attained at about the same wave-length at which the optical absorption becomes high.

The close agreement between the high temperature activation energies and conductance observed in this work and in the work of others

[12] F. Seitz, *Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940), p. 191.

[13] F. Moglick and R. Rompe, Zeits. f. Physik **119**, 472 (1942). J. F. Mullaney, Phys. Rev. **66**, 326 (1944).

on fused silicon indicates that the high resistance values and high temperature coefficients of resistance observed in the films deposited on porcelain are caused by the very high purity of the silicon and not by the crystal size nor by intergranular films. The close agreement between the activation energies obtained from the photoelectric, thermal-conductance, and optical data suggests the phenomena have a common origin —electron transitions from the upper filled band to the conducting band.

Acknowledgment: We are very grateful to Mr. E. F. Kingsbury for assistance in determining the sensitivity of some of the cells and in making observations on the optical absorption of silicon.

The Production of High Centrifugal Fields*

J. W. Beams, J. L. Young, III, and J. W. Moore
Rouss Physical Laboratory, University of Virginia, Charlottesville, Virginia

(Received July 15, 1946)

High centrifugal fields were produced by spinning small solid steel spherical rotors up to their bursting speeds. The rotors were supported magnetically in a vacuum by an improved method and spun by rotating magnetic fields. The peripheral velocities at which the rotors of various sizes, made of the same (flaw free) steel and having the same shapes, exploded, were roughly the same and of the order of 10^5 cm/sec. A centrifugal field of 2.4×10^8 times gravity was obtained with a .795-mm spherical rotor which was the smallest diameter tried. Calculations indicate that in some cases plastic flow probably occurred in small regions near the centers of the spherical rotors somewhat below their bursting speed.

WHEN substances are placed in a centrifugal field two general types of phenomena may occur. The first type depends upon $(2\pi N)^2r^2=v^2$ where N is the number of revolutions per sec., r is the radius of the rotor, and v is the peripheral speed or upon the energy difference between axis and periphery. The second type depends upon $(2\pi N)^2r$ or the intensity of the centrifugal field. In this paper a method is given for producing very high centrifugal fields for use in the study of the second type of phenomena.

In a spinning homogenous elastic rotor the maximum stress produced is proportional to $(2\pi N)^2r^2$ or the square of the peripheral speed.[1] That is, similarly shaped rotors made of the same material will explode when their peripheral velocities are approximately equal. Consequently for the production of the highest centrifugal fields which depend upon $(2\pi N)^2r$ the radius of the rotor must be as small as possible.

For rotors made of a given elastic material the highest peripheral speed should be obtained when the shape is such that the radial and tangential stresses are of constant value throughout the rotor.[1] If hardened high strength steel is used for the rotor, some considerable pains are

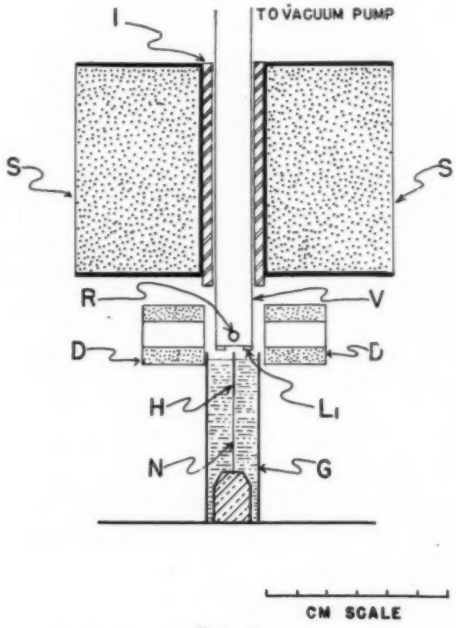


FIG. 1.

* The work described in this paper has been supported by the Bureau of Ordnance, U. S. Navy under Contract NOrd-7873.

[1] Stodola, *Steam and Gas Turbines* (McGraw-Hill Book Company, Inc., New York), Vol. I, p. 376.

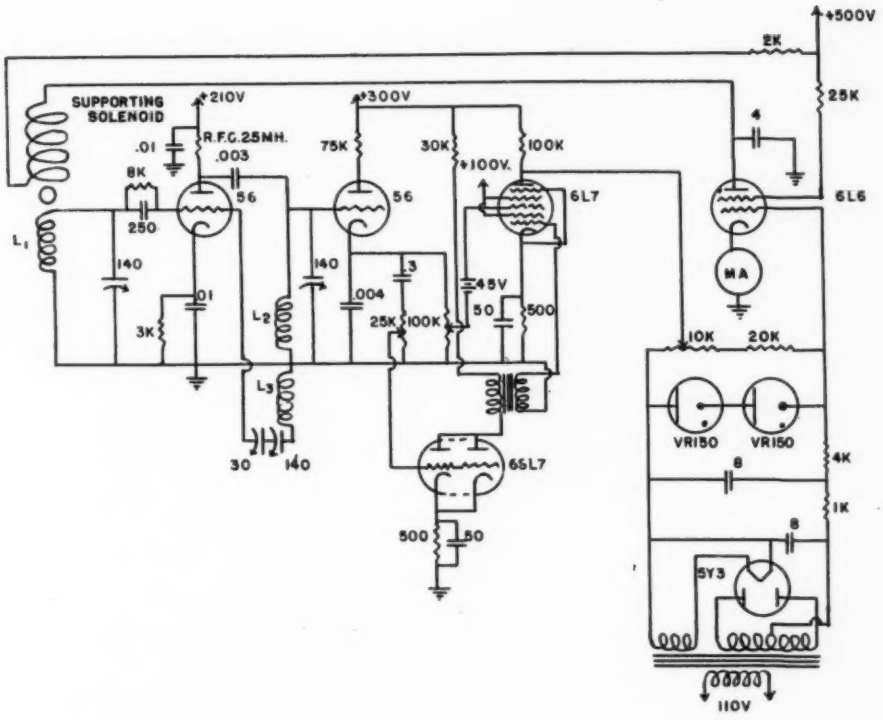


FIG. 2.

required to grind the rotor to the shape required by the above condition. However, spherical rotors are commercially available in the form of high strength solid steel balls for use in ball bearings. Consequently, these approximately spherical rotors were used in the experiments. Also in addition to their availability the calculated critical vibration frequencies of these spherical rotors are well above their bursting speeds.[2]

EXPERIMENTAL ARRANGEMENT

The rotors were spun by the general method developed at Virginia by a number of different workers[3–6] for spinning small rotors. However, in the present experiments, considerable improvements in the apparatus have been made. Figure 1 shows schematically the arrangement used. The rotor R is suspended in the axial magnetic field of the solenoid S and is spun by a two phase rotating magnetic field in the two pairs of coils D. The horizontal position of the steel rotor is maintained by the symmetrically diverging field of the solenoid while its vertical position is maintained by an automatic regulation of the current through the solenoid S. The small coil L_1 is part of the grid circuit of a tuned grid tuned plate radiofrequency oscillator (Fig. 2) which regulates the current through S. It is so arranged that when the rotor rises the current through S decreases. These current changes are such as to position accurately the rotor without observable hunting. An iron tube I with $\frac{3}{4}''$ O. D. $\frac{3}{32}''$ wall thickness and $3\frac{3}{8}''$ long was placed inside the solenoid to increase the field at R. A small iron wire H mounted in a glass tube G filled with a liquid assisted in damping any horizontal motion of the rotor. The glass chamber V surrounding the rotor R was evacuated by the usual fore pump, diffusion pump cold trap combination. The pressure surrounding the rotor was measured by an ionization gauge.

SUPPORTING CIRCUIT

The circuit used for supporting the rotors is shown in Fig. 2. It will be observed that the

[2] Love, *Mathematical Theory of Elasticity* (Dover Publications, New York), fourth edition, p. 285.
[3] F. T. Holmes, Rev. Sci. Inst. **8**, 444 (1937).
[4] F. T. Holmes and J. W. Beams, Nature **140**, 30 (1937).
[5] C. S. Smith, Rev. Sci. Inst. **12**, 15 (1941).
[6] L. E. MacHattie, Rev. Sci. Inst. **12**, 429 (1941).

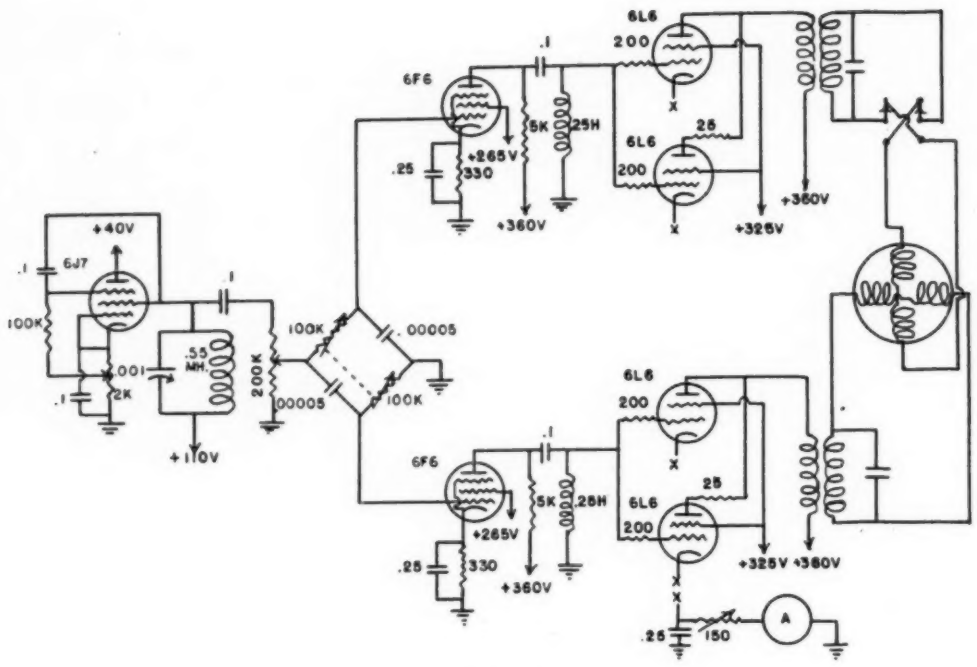


FIG. 3.

"pick-up" coil L_1 is in the grid circuit of the partially neutralized tuned grid tuned plate radiofrequency (10 megacycles which is large in comparison with the rotor speed) oscillator. The effective "Q" of the grid circuit is critically changed by small variations in the height of the rotor R. This change is produced by the reflected impedance due to induced eddy currents in the rotor and to the very slight change in inductance and capacity of the coil L_1. Consider a virtual downward displacement of the rotor. This will reduce the amplitude of oscillation in the grid circuit and, since the oscillator is partially neutralized, the amplitude of oscillation in the plate circuit will also decrease. As a result the potential across the "cathode follower" will likewise drop and cause a lower potential or so-called "error" signal to be applied to the "error" grid of the 6L7 mixer tube. The plate potential of the 6L7 mixer tube will then increase which, in turn, raises the potential on the control grid of the 6L6 power tube. This causes an increase in current through the supporting solenoid which increases the lifting force on the rotor and restores it to its original position. With this arrangement alone small vertical oscillations or hunting of the rotor occurs so additional electrical "damping" was introduced into the circuit. This "damping" is obtained by mixing the amplified derivative of the "error" signal with the "error" signal in the 6L7 mixer tube. The derivative of the "error" signal is obtained from the capacity-resistance circuit in parallel across the cathode follower as shown in Fig. 2. The magnitude of the capacity and of the resistance in this circuit was obtained by observing the hunting frequency of the rotor and then making the capacitative reactance and resistance such that the derivative signal would not be too small and at the same time more than 80° out of phase with the variation of the "error" voltage. The values are given in Fig. 2. This derivative signal is amplified by the two triodes of a 6SL7 in parallel and applied through transformer coupling to the second control grid of the 6L7 mixer tube. This arrangement greatly stabilized the rotor and reduced all vertical hunting and oscillations to such an extent that no vertical motion of the rotor could be observed with a 20-power microscope focused on markings of the rotor. The values of the resistances, capacities and inductances are given in Fig. 2. The supporting solenoid S was wound with 20,500 turns of No. 30 wire. Its inside diameter

was $\frac{13}{16}$" and its length $2\frac{5}{8}$". The coil L_1 was flat wound with 12 turns of No. 32 copper wire. Its effective diameter was about 1.26 cm. L_2 was 2 cm long, 4.3 cm in diameter, and contained 9 turns of No. 14 copper wire. L_3 contained 9 turns of No. 28 copper wire close wound next to the low end of L_2.

DRIVING CIRCUIT

The driving circuit which produces the rotating magnetic field in the coils D which in turn spins the rotor is essentially the same as that used by MacHattie.[6] The 4 coils D are mounted symmetrically around the rotor and opposite pairs are connected in series. The currents in the two pairs of coils differ in phase by approximately 90° and are tuned to the driving frequency by shunting condensers. Figure 3 shows a diagram of the circuit. The output of the transition oscillator is passed through a volume control and then a phase splitting bridge. Each of the phases is amplified and applied to the coils so that the currents in the coils D are 90° out of phase. The frequency used varied from 100,000 cycles per sec. for accelerating the $\frac{5}{32}$" (3.97 mm) rotor to 500,000 cycles/sec. for accelerating the $\frac{1}{32}$" (.795 mm) rotor. The constants given in Fig. 3 were those used to accelerate the $\frac{1}{16}$" (1.59 mm) rotor. It will be observed that the solid steel rotor is actually the armature in an induction motor. In starting the so-called "slip" is very high and gradually decreases as the rotor speeds up. However, the torque at high "slips" gives good acceleration because of the comparative high resistance of the rotor. For example, with the $\frac{1}{16}$" rotor the acceleration was roughly 30 rev./sec.^{2} with approximately $7\frac{1}{2}$ watts input to the coils D. During some of the experiments the coils D were cooled by a small electric fan.

MEASUREMENT OF ROTOR SPEED

At first some difficulty was experienced in getting a simple accurate method of measuring the rotor speed, but the problem was successfully solved by the apparatus shown schematically in Fig. 4. One-half of the rotor is polished and the other half darkened by dipping in dilute H_2SO_4 which had been in contact with metallic antimony. This dark layer is very thin and is not destroyed by the highest centrifugal force so far obtained. Light from an incandescent lamp A is focused on the rotor R and the resulting scattered light focused on the electron multiplier photo-cell P(931A). The variable output from the photo-cell, due to the spin of the rotor, is amplified and fed to one pair of plates of a cathode-ray oscillograph. This frequency of the rotor is then compared in the usual way with the frequency of sweep applied to the other pair of plates of the oscillograph. This sweep frequency was calibrated by a standard frequency oscillator (accuracy from .1 to .01 percent) up to 50,000 cycles per sec. Above this frequency multiples of the calibrated frequencies are used. It is believed that the rotor speeds are precise to three significant figures.

While the above frequency measurements were precise enough for the present experiments, it is highly desirable to have better precision for deceleration measurements. Accordingly, the output of the photo-cell was made to actuate an ordinary scale of 64 counter circuit in such a way that the counter counts the number of r.p.s. directly. This method gives high precision but the present arrangement is not suited for counting above 5×10^4 r.p.s., even with a counter clock, especially wound for the purpose. However, it is hoped to extend its range on up to much higher values.

EXPERIMENTAL RESULTS

As mentioned before the rotors were solid steel spheres (ball bearings). In order to determine the axis of rotation a small flat was ground on each of the rotors. This varied from 1 mil for the $\frac{1}{32}$"

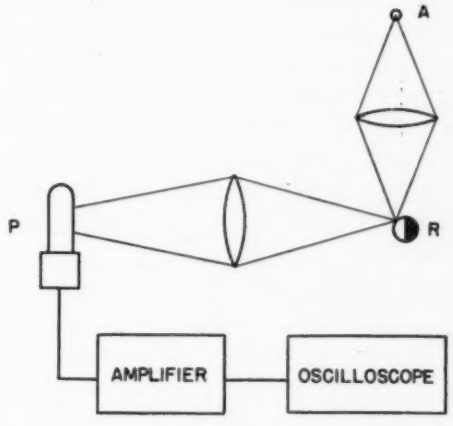


FIG. 4.

TABLE I.

Diam. of rotor mm	Rotor speed r.p.s.	Peripheral speed cm/sec.	Centrifugal acceleration times gravity	Maximum calculated stress lb./in.²
3.97	77,000	9.60×10^{4}	4.71×10^{7}	410,000
2.38	123,500	9.25×10^{4}	7.20×10^{7}	385,000
1.59	211,000	1.05×10^{5}	1.43×10^{8}	498,000
.795	386,000	9.65×10^{4}	2.40×10^{8}	420,000

rotor to 5 mils for the $\frac{5}{32}''$ rotor. This small flat surface became the top of the rotor and hence determined the vertical axis of rotation perpendicular to its surface. The size of the small damping needle H (Fig. 1) together with the viscosity of the liquid surrounding it is determined by the size of the rotor. For a $\frac{1}{16}''$ rotor H was 8 mils and $\frac{1}{2}''$ long, N was 4 mils and $1\frac{1}{2}''$ long and the liquid was "three in one" oil. When the adjustments are properly made no variation in the position of the rotor can be observed. The rotor, the glass vacuum chamber, and the driving coils are surrounded by a wooden barricade which stops the flying parts of the exploding rotor. Although the rotors used were grade A commercial steel ball bearings and supposedly free from flaws, a number of them exploded at abnormally low rotor speeds. These rotors showed definite flaws when their parts were examined after an explosion. However, most of the rotors required very high stresses to explode them. Table I gives some of the results obtained just before the rotors exploded.

It will be observed that the maximum peripheral speed for all of the rotors is approximately 10^5 cm/sec. and determines the bursting speed for the steel used. In accordance with expectations the maximum centrifugal field was obtained with the smallest rotor used and for the .795-mm rotor was 240 million times gravity. In the fifth column are tabulated the maximum stresses in the rotors calculated by the method of Chree.[7] In the calculations it is assumed that the rotors are perfectly elastic. The calculations show that the stresses reach a maximum at the center of the rotor and fall off toward the surface. Actually the steel is well known to be plastic and it is probably that the material in a small region around the center of the rotor flows and relieves the stresses. Therefore, the calculated stresses are probably well above those actually occurring.

[7] C. Chree, Proc. Roy. Soc. **58**, 39 (1895).

The air pressure surrounding the rotor as measured by an ionization gauge was maintained between 10^{-5} and 2×10^{-5} mm of Hg. The "air" friction on the spherical rotor would produce a deceleration, when the rotor is spinning freely with the driving power off, given approximately by

$$\log_e \frac{N}{N_0}=\frac{-5p}{rd}\left(\frac{M}{2\pi RT}\right)^{\frac{1}{2}}(t-t_0),$$

where N_0 is the number of r.p.s. at time t_0; N is the number of r.p.s. at time t; p is the gas pressure in bars, d is the density of the steel = 7.8 g/cm³, T = absolute temperature; R = gas constant, M = molecular weight of the gas, and r = radius of the spherical rotor. With the 1.59-mm rotor spinning freely at about 120,000 r.p.s. with a pressure of about 10^{-5} mm Hg it required roughly two hours to lose one percent of its speed. Therefore, the order of magnitude of this deceleration can be accounted for by gas friction on the rotor. No doubt the rotors are slightly non-homogeneous magnetically and the supporting field is perhaps not strictly symmetrical. Also, in addition, other small factors such as oscillations in the pick-up coil probably add up to give a small residual drag. However, in practice, since the major portion of the drag apparently is due to gas friction, the deceleration can be made very small. With proper coupling between a rotating magnetic field driven by a piezoelectrically controlled oscillator an almost constant speed rotor should be obtained. Consequently many experiments requiring constant rotor speed as well as high centrifugal fields are made possible.

We are much indebted to Drs. L. G. Hoxton and F. Bader for checking some of our calculations.

A Cathode-Ray Tube for Viewing Continuous Patterns

J. B. JOHNSON
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received June 8, 1946)

A cathode-ray tube is described in which the screen of persistent phosphor is laid on a cylindrical portion of the glass. A stationary magnetic field bends the electron beam on to the screen, while rotation of the tube produces the time axis. When the beam is deflected and modulated, a continuous pattern may be viewed on the screen.

IN cathode-ray oscilloscopes as employed for viewing transient phenomena the luminescent screen with persistent after-glow is often very useful, in that the pattern remains on the screen and can be observed in detail for a short time after the excitation. If the pattern is continuous, while varying in content, then the conventional cathode-ray tubes offer no convenient way of observing the trace without interruption. The requirement of continuity arises, for instance, in the application to "visible speech patterns"[1] where the time-pitch-intensity pattern of sounds can be viewed for a few seconds as it is continuously produced. The cathode-ray tube to be described here was designed for experimental use in the speech pattern system and other systems with similar requirements.

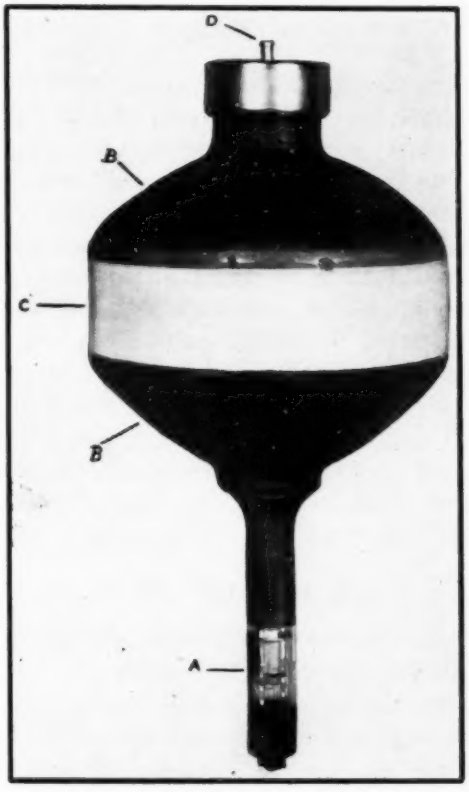


FIG. 1. Photograph of continuous pattern cathode-ray tube. *A*—electron gun; *B*—graphite coating; *C*—luminescent screen; *D*—high voltage terminal.

GENERAL DESCRIPTION

A novel feature of this tube is that the time-axis is produced, not as usual by the electron beam moving across the screen, but by moving the screen horizontally past the beam which remains within a fixed vertical plane, in which plane the deflection takes place. The screen is applied on the cylindrical circumference of the tube, instead of on the end, as shown in the photograph of Fig. 1. The electron gun is contained in the longer neck of the bulb which is coaxial with the screen, and the electron beam is projected along the axis of this neck. By means of a set of coils external to the tube a magnetic field is maintained throughout the large part of the bulb, its direction being perpendicular to the axis of the tube. With suitable strength, this field bends the beam away from the axis within the bulb so that it strikes the screen near its equator. A pair of magnetic deflection coils placed across the narrow neck of the tube permits the beam to be deflected across the screen in a plane through the axis of the tube. Now, with the system of magnetic coils fixed, the tube itself may be rotated about its axis while the beam remains in its fixed plane determined by the stationary bending field. The screen then moves past the phosphorescence-exciting end of

[1] R. K. Potter, "Visible patterns of sound," Science **102**, 463–470, (Nov. 9, 1945). R. R. Riesz and L. Schott, "Visible speech cathode-ray translator," J. Acous. Soc. Am. **18**, 50 (1946).

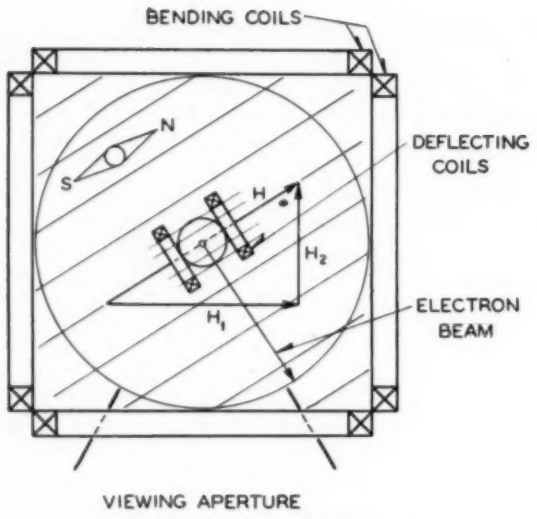


FIG. 2. Horizontal cross section of coil system.

the beam, say from right to left, and a stationary observer sees the history of the trace for the past few moments, the time depending on the rate of rotation. When an exposed area has made a full revolution, the old pattern has become negligibly faint, either by its natural decay-rate alone or with the aid of some wiping-out process such as exposure to light of long wave-length. With these basic features of the tube and its operation made clear, a more detailed description will now be given.

THE BULB AND SCREEN

In order to secure a reasonably long screen the bulb is made large. Its middle portion is a cylinder 10″ in diameter and a little over 3″ in length. From the cylinder the bulb tapers to a wide short neck at one end and a longer and narrower neck for the electron gun at the other. The latter neck is $1\frac{3}{8}$″ in diameter so as to fit in standard magnetic deflection yokes. The over-all length of the based tube is 21 inches.

The screen is deposited on the inside of the cylindrical part of the bulb, 3″ wide. It is of the P7 type of phosphor, with persistent yellow after-glow. The parts of the bulb not occupied by the screen are covered with a layer of colloidal graphite, forming part of the electrode system.

THE ELECTRON GUN

The kind of electron gun is not critical, but electrostatic focusing guns like those of the

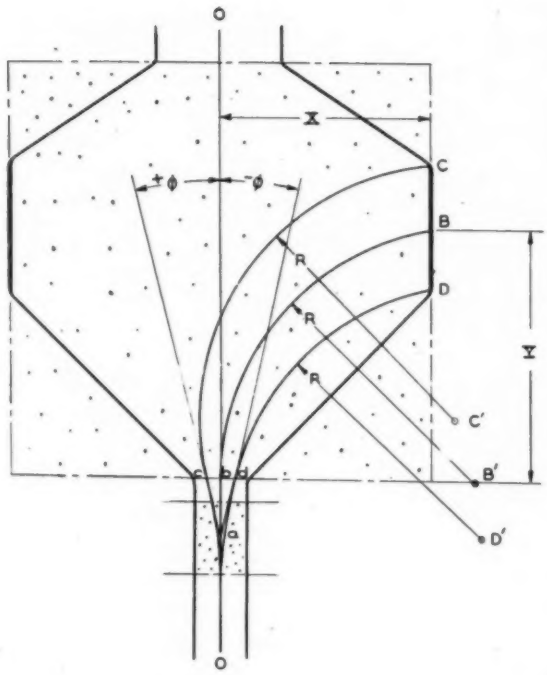


FIG. 3. Vertical cross section of tube in magnetic field. Dots indicate magnetic field directed into the paper. Electron beams *OB*, *OC*, *OD*.

Western Electric 325-A tube or the RCA 7CP tube have been used, the tube of Fig. 1 having the latter type of gun. The electron beam of this gun can be modulated, the cut-off point being at about −45 volts and the maximum beam current several hundred microamperes. The graphite coating on the bulb forms the final electrode of the gun structure, operated at potentials of 5000–7000 volts. This voltage is brought in through the base at the top of the tube, all other voltages being connected through the smaller base at the gun end. The gun must be lined up very accurately with the axis of rotation of the bulb in order to prevent wandering of the spot while the tube turns.

THE MAGNETIC FIELDS

The arrangement of the coils about the tube is indicated in Fig. 2 which presents a top view of the system. The bending field is produced by a set of four roughly square coils whose sides about equal the diameter of the tube. By using four coils the viewing area can be centered through one of them as shown, with the beam to one side of the observer. A field strength of about 18

oersteds is required for bending the beam to the center of the screen. This field, in a suitable direction, is produced by a mmf of about 640 ampere-turns in each of the side coils and 400 ampere-turns each in the front and back coils, the ratio of field components being $H_1:H_2=8:5$. This field is not very uniform but since the beam makes only a small vertical excursion in it the uniformity is not of great moment.

Below the bending coils and outside their main field is mounted the deflection yoke, shown in the illustration as a pair of air-core coils embracing the neck of the tube above the electron gun. This yoke must be aligned so that its magnetic field is closely parallel to the effective direction of the bending field, for if the fields are not parallel the beam is deflected in an S-shaped pattern instead of a straight vertical line. Distortion is also produced if the deflecting field overlaps appreciably the focusing fields of the electron gun. The deflection sensitivity is of the order of 1 inch at the screen per 50 ampere-turns in the deflection coils.

THE ELECTRON PATHS

In Fig. 3 is shown a cross section of the tube, in the plane of the electron beam, perpendicular to the magnetic fields. For simplicity it is assumed that the magnetic fields are uniform and sharply bounded at the limits of the coil systems, as indicated by dotted areas. The outlines of the tube are shown to scale, with the axis 0–0 and the screen at *C–D*. Let the radius of the screen be X and the distance from the beginning of the bending field to the center of the screen at B be Y. When the undeflected beam is injected into the bending field it describes the path *abB* if the field is adjusted so that the radius of curvature of the beam is

$$R=(X^2+Y^2)/2X.$$

The field is adjusted for this position of the undeflected beam. When now the beam is given a small deflection it enters the bending field as if coming in a straight line from the center point a of the deflecting field, making the angle φ with the axis, and then proceeds in a circle of the same radius R, tangent to the line of entrance, until it strikes the screen. Extreme rays are shown as *acC* and *adD*.

On the basis of the idealized field distribution the deflection of the beam from the center of the screen is given very closely by the expression

$$\Delta=Y-y=R\left(\frac{2X}{R}-\frac{X^2}{R^2}\right)^{\frac{1}{2}}$$

$$-R\left(\frac{2X}{R}\cos\varphi-\frac{X^2}{R^2}+\sin^2\varphi\right)^{\frac{1}{2}}+R\sin\varphi,$$

where X is the radius of the screen. The middle term in this expression is of the second order in φ so that for small angles of deflection the first two terms nearly cancel and

$$\Delta\doteq R\sin\varphi\doteq R\varphi.$$

Within the range of the screen the deflection is surprisingly linear, in spite of considerable variation in the length of the beam. Astigmatism and defocusing would be expected with the type of path these beams follow, but neither of these has been found very serious.

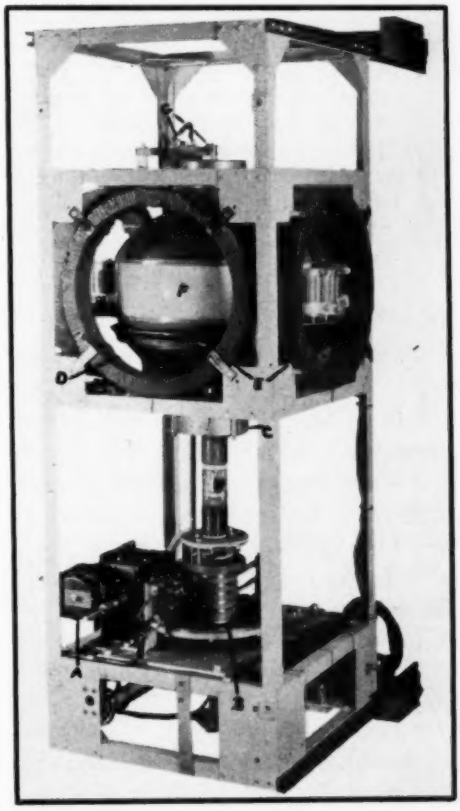


FIG. 4. Mounting rack for the tube. *A*—Motor; *B*—slip-rings; *C*—deflection yoke; *D*—driven carriage; *E*—bending coils; *F*—screen; *G*—high voltage terminal.

MOUNTING OF THE TUBE

Two methods have been used for supporting and driving the tube. In the first, the tube is mounted by the bases at each end and driven through the lower base. The preferable method is to support the main body of the tube in a driven cradle, with the bases used only for the electrical connections. In either case all connections are made through slip-rings which are driven with the tube. The rate of drive has been in the range of one to twelve revolutions per minute, or .5 to 6 inches per second at the screen. Figure 4 shows the tube mounted in a frame with bending and deflection coils and with drive of the second kind.

ACKNOWLEDGMENT

The building of a tube of this kind presented many new problems of structure and procedure, in the solution of which Mr. J. B. Little, Mr. D. A. S. Hale, and Mr. C. L. Loudon deserve much credit.

Improvements in the Stability of the FP-54 Electrometer Tube*

J. M. LAFFERTY AND K. H. KINGDON
General Electric Company, Schenectady, New York

(Received April 23, 1946)

It was found that a considerable portion of the instability in the FP-54 is caused by variations in emission from the thoriated tungsten filament. Operation of the filament at a current which neither activates nor deactivates was found to be a good criterion for adjusting the circuits employing these tubes. Large fluctuations were observed in emission immediately after activation of the filament. Greater stability was obtained by increasing the activation time from 8 to 40 minutes. Filament end shields improved the stability, showing that some rapid fluctuations in emission occur at the poorly activated end portions of the filaments. Long-time drifts were not improved by end shields. Tubes with oxide-coated filaments gave greater sensitivity and less grid current than tubes with thoriated tungsten filaments. These tubes, however, had a tendency to drift. A split type of FP-54 was constructed with a common filament and space-charge grid, but with twin control grids and plates. Both oxide-coated and thoriated tungsten filaments were used. Operation of these tubes in a bridge type circuit eliminated long-time drift and decreased the amplitude of rapid fluctuations.

THE FP-54 pliotron is a low grid current vacuum tube commonly used as an electrometer for measuring small currents. This tube has been described by Metcalf and Thompson.[1]

The sensitivity which can be utilized with this tube and its associated circuit using a sensitive galvanometer is limited by short-period fluctuations and long-period drifts. It is the purpose of this paper to discuss the results of investigations made to determine the cause of these variations and methods of correcting for them.

CAUSES OF INSTABILITY

The causes of instability in the FP-54 tube and its associated circuit can be listed as follows:

1. External varying electric fields.
2. External varying magnetic fields.
3. Irregular leakage over the surface of the tube envelope.
4. Loose connections and poor rheostat contacts.
5. Microphonics.
6. Variations in supply voltage.
7. Variations in filament emission.

Figure 1 shows a diagram of the circuit used in determining instability. With switch S_2 closed, stability was measured by reading the galvanometer deflection every half-minute for a period of one-half hour and plotting a curve of galvanometer deflections against time. The 1.5-volt cell, with R_1 and R_2, forms a potentiometer arrangement used in conjunction with S_2 for increasing the grid bias 10 millivolts when making sensitivity measurements on the tube.

The following precautions were taken to elimi-

* An abstract of the material presented in this paper was given at the meeting of the American Physical Society in Cambridge on April 27, 1946.

[1] G. F. Metcalf and B. J. Thompson, Phys. Rev. **36**, 1489 (1930).

nate or minimize the causes of instability listed above:

1. Tube apparatus, including all wiring, batteries, galvanometer, and shunt were operated in shielded compartments connected by shielded cables.
2. It was not found necessary to shield the tube and circuit from magnetic fields.
3. The envelopes of all tubes tested were washed with alcohol and painted with Aquadag to within $\frac{3}{4}$ inch of the control grid connection. This Aquadag conductor was then grounded at a potential equal to that of the control grid. The tube was operated in an air-tight grounded brass chamber containing phosphorous pentoxide. A rough vacuum was maintained in the chamber during tests.
4. Large heavy duty wire wound resistors, rheostats, and potentiometers were used throughout the circuit. All connections were soldered and made mechanically rigid.
5. The tube was mounted on sponge rubber, and electrical connections to it were made through flexible leads.
6. Three heavy duty Philco storage batteries were used. They are of the high gravity type.
7. Effects due to variations in supply voltage and filament emission will be discussed in detail below.

After the precautions noted above have been taken, the most important remaining causes of instability are fluctuations in storage battery voltage and fluctuations due to random discontinuities in filament emission. Figure 1 is one of the modified circuits of DuBridge and Brown.[2] The resistances in this circuit may be adjusted in such a manner that the galvanometer reading will be independent of small changes in battery voltage. Further, by balancing the space-charge grid current against the plate current, the galvanometer will remain unaffected by fluctuations in filament emission which change the space-charge grid and plate currents in the *same ratio.*

If for the circuit of Fig. 1, the galvanometer deflection is plotted against the battery current,

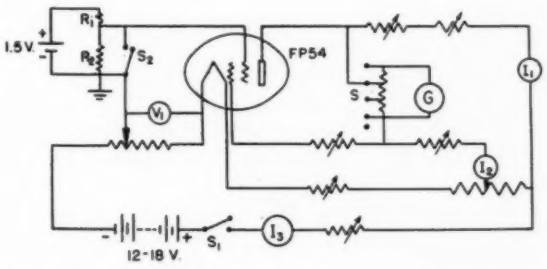


FIG. 1. Diagram of circuit used for testing FP-54 pliotrons.

[2] L. A. DuBridge and H. Brown, Rev. Sci. Inst. **4**, 532 (1933). D. B. Penick, Rev. Sci. Inst. **6**, 115 (1935).

there will, in general, be found two balance points, marked X and Y in Fig. 2, where $\partial d/\partial I_3=0$. If the circuit is operated at point X, one would expect any changes in battery current to cause a decrease in the galvanometer reading. However, cases were observed where the galvanometer reading actually increased. Cases were also observed where the galvanometer reading decreased when the circuit was operated at point Y. These galvanometer deflections might

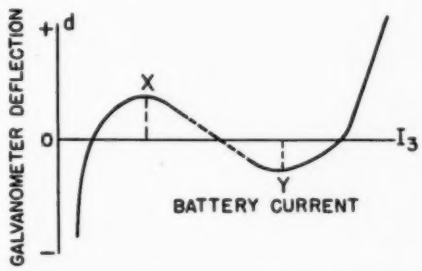


FIG. 2. By proper adjustment of the resistances in Fig. 1, balance points X and Y may be found where the galvanometer deflection is independent of small changes in battery voltage.

have been caused by increases or decreases in filament emission which did not change the space-charge grid and plate currents in the same ratio, but could not have been caused by battery current changes. It was found that for the batteries used, the galvanometer fluctuations and drift were no worse when the circuit was operated off a balance point. A much better criterion for adjusting the circuit was to operate the filament at a current which neither activates nor deactivates it. This is, in general, between 80 and 95 milliamperes.

ACTIVATION EFFECTS

Thoriated tungsten filaments are usually given the following treatment to make them good electron emitters. First, the filament is heated to 2890°K for a short period of time to clean the tungsten surface and reduce some of the thorium oxide in the tungsten. Any thorium appearing on the surface is immediately evaporated at this temperature. This operation is called flashing. Second, the filament is heated to 2155°K for several minutes. During this activation process metallic thorium diffuses to the surface and a portion of it evaporates producing a surface

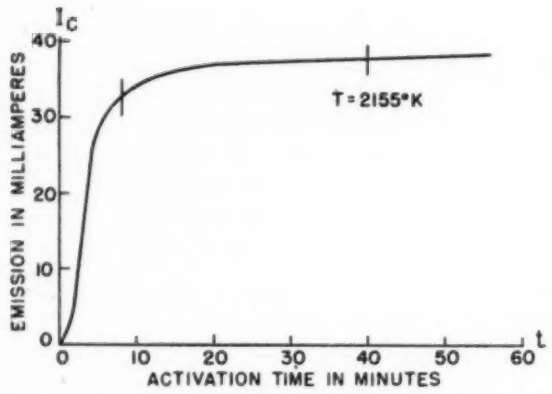


FIG. 3. Filament emission as a function of activation time for the FP-54.

layer of molecular thickness. The thorium layer produced at this temperature gives maximum emission. Third, the filament is reduced to 1800°K for normal operation. However, in the case of the FP-54 the filament is operated at 1650°K to reduce photoelectric grid emission.

For filaments which were just activated, very large and rapid galvanometer fluctuations (100 to 150 millimeters with approximate periods of $\frac{1}{3}$ second) were observed, indicating spasmodic changes in emission. These effects decreased with time, and after the tubes had been in operation for 24 to 48 hours, they behaved in a more normal fashion.

Figure 3 shows how the filament emission increases with the activation time. Increasing the activation time beyond eight minutes does not increase the emission appreciably. However, it was found that increasing the activation time to 40 minutes gave a marked increase in stability. This is shown by the curves in Fig. 4. A stability curve was first run on a tube which had been activated for the usual eight minutes. The filament of this tube was then flashed at 250 milliamperes for one minute and activated at 150 milliamperes for 40 minutes. A stability curve was then run again after the tube had been in operation for 24 hours. It can be seen that the stability is improved and the sensitivity increased slightly.

Tests were also made on tubes with shields over the ends of the filament (described below), and it was found that the stability improved, if the filaments were activated for the longer period of time.

These facts would indicate that the longer activation time not only activates the cooler end

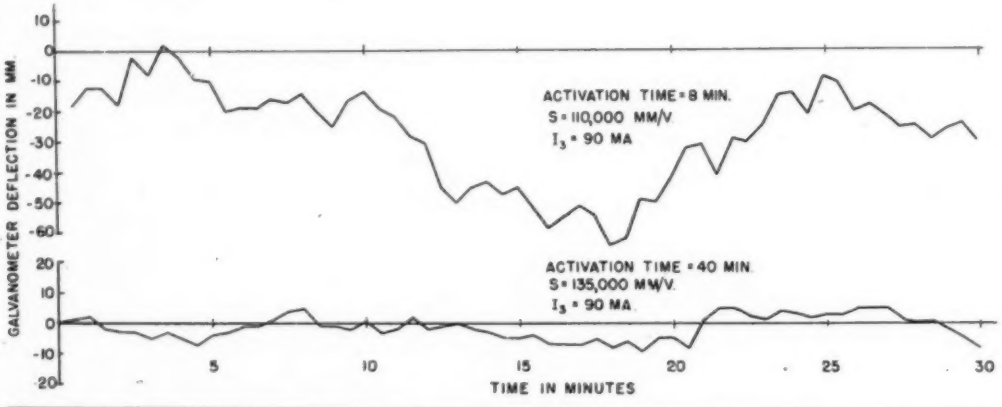

GALVANOMETER DEFLECTION IN MM.
ACTIVATION TIME = 8 MIN.
S = 110,000 MM/V
I_3 = 90 MA
ACTIVATION TIME = 40 MIN.
S = 135,000 MM/V
I_3 = 90 MA
TIME IN MINUTES

FIG. 4. Effects of increasing the activation time on the stability of the FP-54.

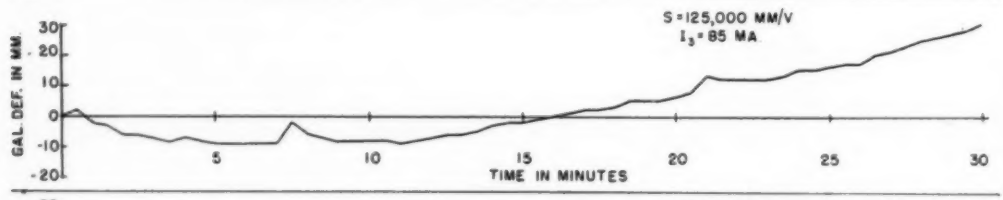


FIG. 5. Stability curve for an FP-54 with filament end shields.

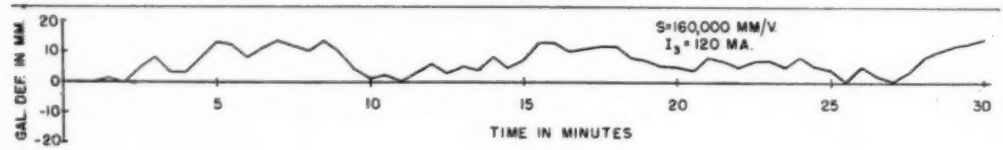


FIG. 6. Stability curve for an FP-54 with an oxide-coated filament.

portions of the filament more completely, but also establishes a more stable equilibrium between the thorium on and below the surface throughout the entire length of the filament.

FILAMENT END EFFECTS

If the instability of the FP-54 is caused by spasmodic changes in filament emission, one would reasonably assume that the effects would be more pronounced near the ends of the filament, where it operates temperature-limited, than along the center portion, where it operates space-charge limited.

Some tubes were constructed, having shields around the ends of the filaments, so that only emission from the hot middle section reached the grids and plate. The shields were connected to the negative side of the filament lead. Figure 5 shows a typical stability curve for this type of tube. Some of the small fluctuations are smoothed out, but the tube still has a tendency to drift over a long period of time. The sensitivity is reduced slightly because part of the active portion of the filament is shielded.

STABILITY OF OXIDE-COATED FILAMENTS

Some FP-54 tubes were constructed, using oxide-coated filaments in place of the thoriated tungsten. The filaments used were of the directly heated type, consisting of a coated one-mil nickel alloy wire 1.5 inches long. These filaments had a rating of 1.5 volts at 120 milliamperes.

A much higher emission was obtained with the oxide-coated filaments operating at a dull red heat (700°C) than with the thoriated tungsten filaments operating at 1374°C. The space-charge-limited emission was also higher for a given anode voltage, because there was less voltage drop across the oxide-coated filament. The increased emission gave the tube a higher mutual conductance and hence a higher sensitivity.

Figure 6 shows a typical stability curve for an FP-54 with an oxide-coated filament. In general, these tubes have a tendency to drift. It was found that if any of the electrode voltages were increased or decreased slightly, the emission would increase or decrease slowly for two or three hours after the change had been made. This effect is very undesirable, since it causes the galvanometer zero to drift.

The grid current for these tubes was about 5×10^{-16} ampere. This is about one-tenth the grid current for the standard tube with the thoriated tungsten filament. This reduction in grid current is undoubtedly caused by the reduction in photoelectrons emitted by the control grid under the action of light from the filament.

SPLIT ELECTROMETER TUBE

The use of two electrometer tubes in a Wheatstone bridge arrangement as a means of compensating for fluctuations in anode voltage and filament current has been described in the literature.[3] Because of the symmetrical arrangement, the effect of external disturbances on the grid circuits can also be compensated for.[4] Random fluctuations in emission from the individual filaments are uncompensated, however.

DuBridge and Brown[2] have attempted to compensate for fluctuations and drift in filament emission by balancing the plate current against the space-charge grid current. Once this bridge arrangement is balanced, the galvanometer current will remain zero for cases where drift and fluctuations in emission change the plate and space-charge grid currents in the same ratio. This circuit, however, is handicapped by the asymmetry of the space-charge grid and plate. The space-charge grid draws about four or five times as much current as the plate. Under these conditions it was found that drift and fluctuations in plate and space-charge grid current do not change in the same ratio and that emission fluctuations are not entirely compensated for.

Brentano and Ingleby[5] have suggested that the advantages of the two-tube bridge circuit could be retained, and in addition random filament current fluctuations could be compensated for, if the two tubes were replaced by a single tube with a common filament supplying current to two separate sets of grids and plates. They have described some preliminary experiments showing the feasibility of this method. This principle as applied to the FP-54 structure is described in the experiments below.[6]

[3] J. C. M. Brentano, Nature **108**, 532 (1921). L. A. DuBridge, Phys. Rev. **37**, 392 (1931).
[4] J. C. M. Brentano, Phil. Mag. **7**, 685 (1929).
[5] J. C. M. Brentano and P. Ingleby, J. Sci. Inst. **16**, 81 (1939).
[6] At the time these experiments were made, in the summer of 1940, we did not know of the work of Brentano and Ingleby.

Computations[7] show that the potential minimum around the filament of the FP-54 is located out from it at a distance of at least 12 diameters. This distance should be great enough to allow the space-charge to redistribute itself almost uniformly around the filament if a change in emission occurred on one side of the filament. Thus, changes in emission current should be nearly the same in all directions. The larger the space-charge density, the more nearly true this

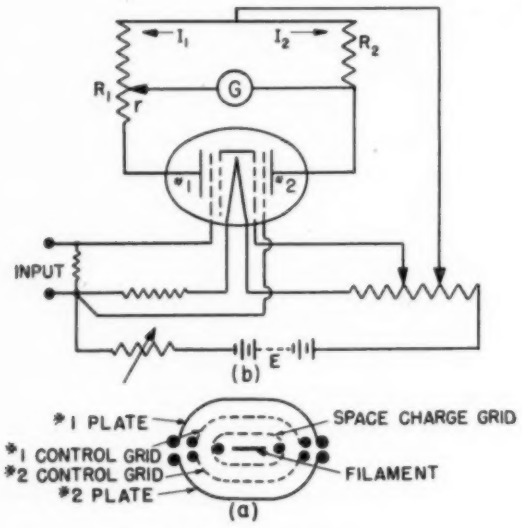


FIG. 7. Basic circuit diagram and electrode structure for the split FP-54.

should be. It was thus decided to modify the FP-54 structure as shown in Fig. 7a, cutting the control grid and plate in two sections. Each section is placed in the arm of a Wheatstone bridge as shown in Fig. 7b.

Assuming for the moment that each section has identical characteristics, a change in filament emission would cause the plate current in each arm of the bridge to change by equal amounts. Thus, if the bridge is balanced, it will remain so independent of changes in filament emission. The input voltage impressed across the No. 1 grid upsets the bridge balance and gives a galvanometer deflection.

For maximum sensitivity, the resistors in the bridge arms should have the same value as the plate resistance of one section of the tube. By using an FP-54 tube split in half, the sensitivity will be decreased to one-quarter of that obtained for a single tube circuit with the galvanometer in the plate circuit. This is because the mutual conductance is reduced to half-value by splitting the tube, and the bridge arrangement decreases the sensitivity by a factor of one-half.[8]

If the split tube could be built exactly symmetrical, so that each half had identical characteristics, the bridge balance would be independent of variations in the supply voltage E. Actually, such tubes cannot be constructed, and circuits must be used which will compensate for differences in characteristics.

The circuit conditions to be satisfied in Fig. 7 are

$$I_1 = I_2, \tag{1}$$

$$\partial I_1/\partial E = \partial I_2/\partial E, \tag{2}$$

and

$$R_1 - r = R_2. \tag{3}$$

Because of symmetry, it was felt that (1) would be the best condition for balancing out drift and fluctuations in filament emission. Condition (2)

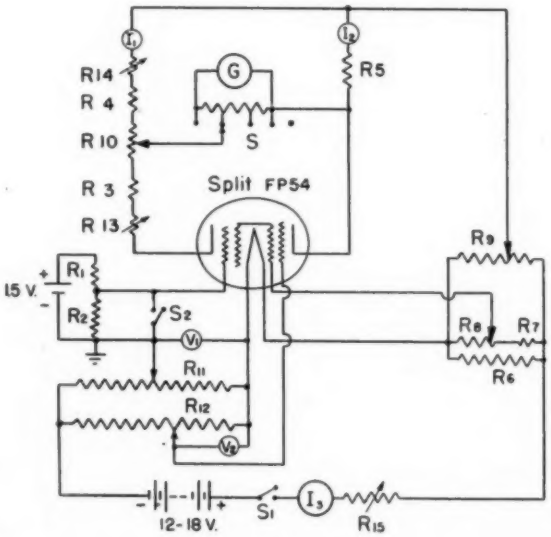


FIG. 8. Split FP-54 circuit diagram.

R_1 = 150,000 ohms
R_2 = 1000 ohms
R_3 = 0–20,000 ohms
R_4 = 0–20,000 ohms
R_5 = 50,000 ohms
R_6 = 100 ohms
R_7 = 75 ohms
R_8 = 125 ohms
R_9 = 200 ohms
R_{10} = 5000 ohms
R_{11} = 125 ohms
R_{12} = 125 ohms
R_{13} = 400 ohms
R_{14} = 50,000 ohms
R_{15} = 35 ohms
S = Ayrton shunt
S_1 = knife switch
S_2 = knife switch
V_1 = 0–10 voltmeter
V_2 = 0–10 voltmeter
I_1 = 0–100μ ammeter
I_2 = 0–100μ ammeter
I_3 = 0–100 milliammeter
G = galvanometer (2×10^{-10} amp./mm)

[7] I. Langmuir, Phys. Rev. **21**, 419 (1923).

[8] W. Nottingham, J. Frank. Inst. **209**, 287 (1930).

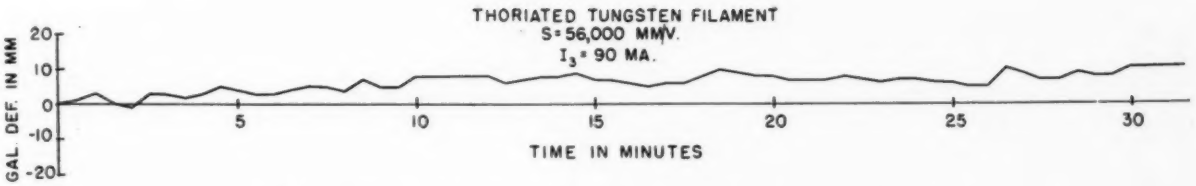


FIG. 9. Stability curve for a split FP-54 with a thoriated tungsten filament.

must be satisfied if the bridge is to remain balanced, independent of small variations in battery voltage. The requirement for bridge balance is given by (3), assuming (1) is satisfied.

Figure 8 shows a diagram of the actual circuit used with the split FP-54 tube. The filament current is adjusted to its rated value by means of R_{15}. R_8 and R_9 are used to adjust the space-charge grid and plate voltages, respectively. The grid biases are adjusted by means of R_{11} and R_{12}. The method of balancing this circuit so that conditions (1)–(3) are satisfied is as follows: With the tube operating at normal voltages, the grids are biased equally at -4 volt. The Ayrton shunt is placed on the ∞ tap, and equal resistances are placed in each arm of the bridge. The bias V_1, on the control grid, is kept constant at -4 volts and V_2 is adjusted until $I_1 = I_2$. R_{15} is decreased until I_3 is increased 10 to 15 percent, and I_1 and I_2 are observed. If I_1 is greater than I_2, R_{14} is increased; or if it is less, R_{14} is decreased. I_3 is then reduced to its normal value and V_2 is again adjusted until I_1 equals I_2. These adjustments are repeated until I_1 and I_2 track, i.e., remain equal for a 10 to 15 percent change in I_3.

The galvanometer is now connected in the circuit. It may be necessary to insert a resistance R_3 or R_4, to obtain a balance within the range of R_{10}, but the total resistance of the left arm must be kept constant. With the circuit balanced, the galvanometer is observed for a time to determine if there is any drift. If drift is present in a direction that would indicate I_1 to be increasing more rapidly than I_2, then R_{13} or R_{14} must be increased slightly. R_{13} gives a fine adjustment of R_{14}. If I_1 appears to be decreasing more rapidly than I_2, then R_{13} or R_{14} must be decreased slightly. By the proper adjustment of R_{13} or R_{14}, galvanometer drift can be entirely eliminated.

Figure 9 shows a typical stability curve for the split FP-54 with a thoriated tungsten filament. A picture of the tube is shown in Fig. 10. It is constructed from standard FP-54 parts. The control grid and plate have the same dimensions as those of the standard tube, the only difference being that they are cut in two sections. One section of the control grid is supported on the quartz insulators, and the lead is brought out through the top of the envelope. No special precautions were taken to prevent leakage current to the other section of the control grid. An octal base was used for this tube. The standard FP-54 uses a four pin base.

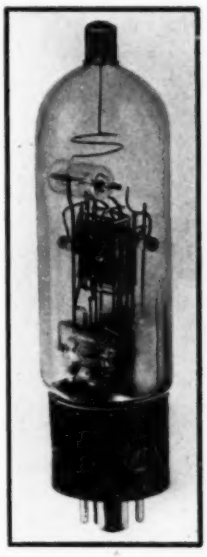


FIG. 10. The split FP-54.

Encouraged by the high sensitivity and low grid current obtained with the oxide-coated filament in the single FP-54 structure, it was decided to try this filament in the split tube. Figure 11 shows a typical stability curve obtained. The sensitivity has been increased to 70,000 millimeters per volt. The objectional drift mentioned above for the oxide-coated filament was entirely compensated for in the split tube. Table I gives the operating conditions

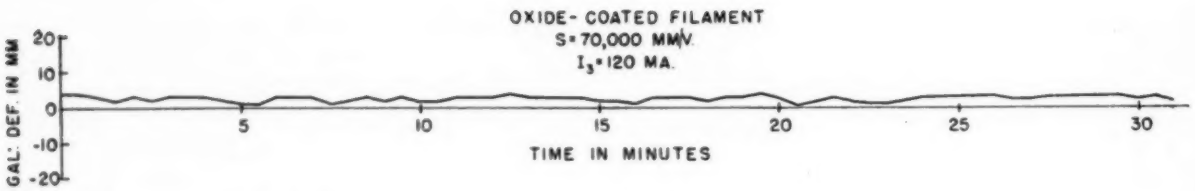


FIG. 11. Stability curve for a split FP-54 with an oxide-coated cathode.

TABLE I. Split FP-54 characteristics.

Filament voltage	1.5 volts
Filament current	0.12 amp.
Space-charge grid voltage	+4.0 volts
Control grid voltage	−4.0 volts
Plate voltage	+6 volts
Mutual conductance	25 μa/volt
Plate resistance (per plate)	40,000 ohms
Plate current (per plate)	60 μa
Control grid current	10^{-15} amp.
Control grid capacity	6.5 μμf

and tube characteristics for the split tube with an oxide-coated cathode.

This tube performed very well indeed, and its stability can be depended upon day after day. Dr. Apker of our laboratory has been using one of these tubes in connection with his photoelectric investigations. His circuit has a sensitivity of 50,000 millimeters per volt and uses a grid resistor of 10^{10} ohms. He reports that the long-time drift is less than 10 centimeters per week and that the amplitude of the short-time fluctuations is about three times greater than the thermal noise of the grid resistor, computed for a band width of one cycle per second.

Calendar of Meetings

November 1946

7–9 Conference on Electrical Insulation, Division of Engineering, and Industrial Research, National Research Council, Baltimore, Maryland
14–16 Acoustical Society of America, Chicago, Illinois
29–30 American Physical Society, Minneapolis, Minnesota

December 1946

2–6 American Society of Mechanical Engineers, New York, New York
5–7 Electron Microscope Society of America and the American Society for X-Ray and Electron Diffraction, Pittsburgh, Pennsylvania (Joint Meeting)
17 Institute of Aeronautical Sciences, New York, New York
26–28 Geological Society of America, Chicago, Illinois

January 1947

3–4 American Physical Society, Los Angeles, California
15–17 American Society of Civil Engineers, New York, New York
23–25 American Association of Physics Teachers, New York, New York
23–25 American Physical Society, New York, New York

February 1947

20–22 Optical Society of America, New York, New York
24–28 American Society for Testing Materials, Philadelphia, Pennsylvania

A Picturesque Oxidation of Molybdenum Vacuum-Tube Leads

HERMAN A. LIEBHAFSKY
Research Laboratory, General Electric Company, Schenectady, New York
(Received July 10, 1946)

A picturesque formation of MoO_3 on vacuum-tube leads gains in interest and importance because of its close resemblance to a growth of W_4O_{11} on tungsten powder in the electron microscope.

DURING the annealing of certain large vacuum tubes in nitrogen from the factory supply, some oxidation of the six molybdenum leads is always possible. Very rarely this oxidation results in a luxuriant growth of white needles like that in Fig. 1. An almost exact counterpart of this phenomenon on a much reduced scale has been observed by Schmidt[1] on tungsten powder in the electron microscope.

The nitrogen is known to contain oxygen and water vapor, each at a variable pressure usually near several millimeters. The gas enters the lower section of the annealer, which houses the photographed portion of the tube and is separated from the upper by an asbestos partition. The lower section is maintained near 400°C; in the upper, where most of the annealing occurs, the temperature is regulated as follows: (1) raised from room temperature to 580° in 45 min. (2) kept at 580° for 30 min. (3) lowered to 400° in 90 min., whereupon the tube is removed. The ceramic spacer (see photograph) keeps the leads aligned during the heating; since this particular spacer had served many tubes, there is no reason to regard it as a source of oxidizing agent.

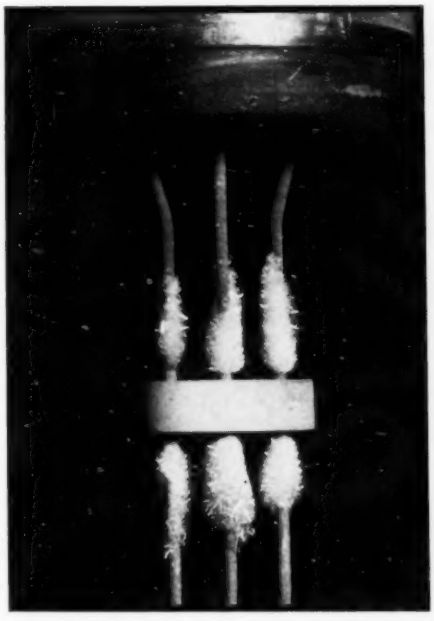


FIG. 1. The oxidation of molybdenum vacuum-tube leads (diameter, $\frac{1}{8}$ inch).

Qualitative analysis indicated, and x-ray diffraction work proved, that the crystals were MoO_3. A few simple experiments intended to produce them at will failed altogether; a stunted yellow growth covering the heated surface was all that resulted—apparently Schmidt's work on molybdenum powder in the electron microscope[1] turned out the same way.

Schmidt concluded, by comparing electronmicrographs, that his crystals were W_4O_{11}. The formation, in relatively large amount and in direct contact with the metal, of the highest oxide, but one, of tungsten and of the highest oxide of molybdenum is a striking result—especially in view of the low pressures of the oxidizing agents and of the virtually complete absence of lower oxides. The lower oxides normally form at the surface when a metal is oxidized. It consequently seems necessary to assume that the MoO_3 and the W_4O_{11} are formed above this surface by the action of the oxidizing agent on the lower oxides; oxygen is much more likely to accomplish this than is water vapor.

With neither metal is it possible to predict the occurrence of the pronounced crystal growth at a definite point. It consequently seems plausible that an exceptional condition ("active spot") is required to initiate this growth, which suggests the following picture: At an active spot, oxidation of the metal is rapid enough to produce an

[1] R. W. Schmidt, Kolloid Zeits. **102**, 15 (1943).

unusually porous structure at the surface. Oxygen continues to diffuse through this structure to generate lower oxides, the oxide superstructure being pushed out into space, where it reacts with oxygen to form the highest oxide (MoO_3 or W_4O_{11}) that is stable under the experimental conditions.

Let us interpret the photograph with this picture in mind. There must be a temperature gradient along the leads during annealing; except near the spacer, the flow of gas is predominantly upward, both because the nitrogen is introduced at the bottom, and because of convection currents. The salient features of the crystal growth with suggested explanations in parentheses are: (1) its lower limit (temperature too low further down) (2) clumps of crystals usually broad above base and narrowest at top (progressively less oxygen as top is approached; note, however, that some crystals have been lost) (2) lower clumps do not quite reach spacer (oxygen "starvation" and physical interference of spacer) (4) upper clumps begin some distance above spacer (dead space in a gas stream) (5) upper limit (temperature too high and/or oxygen "starvation").

Evaporation and condensation have been ignored thus far, even though MoO_3 and W_4O_{11} are both volatile. Schmidt's electron-micrographs (especially Fig. 1) show thin, long, isolated crystals nearly perpendicular to the metal surface. These crystals grow in a rapidly moving gas stream at a pressure near 10^{-5} mm. The mean free path being large, they could scarcely have grown by condensation. Evaporation and condensation are more likely to have been at work in the *highly branched* growth that Schmidt (Fig. 4) observed on a substratum of W_4O_{11}, although chemical factors are not completely ruled out even here.

I wish to thank Mr. T. G. Crawford, Electronics Department, for calling the crystals to my attention, and Dr. David Harker for supervising the x-ray work that proved their identity.

Luminescence of $(Zn, Be)_2SiO_4$:Mn and Other Manganese-Activated Phosphors

JAMES H. SCHULMAN*
Sylvania Electric Products, Inc., Salem, Massachusetts
(Received June 10, 1946)

Spectral energy distribution and luminous efficiency measurements of Zn_2SiO_4:Mn and of various samples of $(Zn, Be)_2SiO_4$:Mn are reported. Quantum efficiencies of these phosphors under 2537A excitation are calculated from this data. The introduction of beryllium into alpha-willemite causes a decrease of the "green" luminescence band of this phosphor and the growth of a new emission band with a peak in the neighborhood of 6100A, the high quantum efficiency being preserved. The general role of manganese as an activator is considered, and the resulting picture used to explain the effect of beryllium and other substituents for Zn^{+2} or Si^{+4} on the spectral properties and on the luminescence decay of manganese-activated materials.

IT is well known that the photo-luminescence color of $(Zn, Be)_2SiO_4$:Mn phosphors may be varied over a wide range, from green to red, by varying the Zn:Be ratio and the concentration of the manganese activator. Spectral energy distribution data corresponding to these variations are very meager. According to Leverenz and Seitz[1] substitution of zinc by beryllium "flattens the [5250A-peaked] alpha-willemite spectrum and shifts it to the red." The measurements of Kröger[2] on two samples of zinc beryl-

TABLE I. Composition of phosphor samples.

Sample number	Moles ZnO	BeO	MnO	SiO_2
A	2.0	0.0	0.122	1.18
B	1.8	0.2	0.058	1.11
C	1.8	0.2	0.081	1.11
D	1.8	0.2	0.111	1.17
E	1.8	0.2	0.138	1.20
F	1.6	0.4	0.132	1.19

* Present address: Crystal Section, Naval Research Laboratory, Washington, D. C.
[1] H. W. Leverenz and F. Seitz, J. App. Phys. **10**, 479 (1939).
[2] F. A. Kröger, Physica **6**, 764 (1939).

TABLE II.

Sample number	Efficiency of luminescence spectrum (lumens per watt)	Quantum ratio	Corrected for: Conversion of electrical energy to 2537A	Corrected for: Unabsorbed 2537; absorption of luminescent light by bulb, phosphor, and lamp ends	Observed efficiency (−5 lumen per watt)	Quantum efficiency
A	517	242	145	124	95	0.77
B	402	174	104	89	79	0.89
C	365	155	93	80	71	0.89
D	345	145	87	75	63	0.84
E	323	135	81	69	58	0.84
F	278	114	68	58	49	0.84

lium silicate at room temperature indicate, on the other hand, that the substitution calls forth a new emission band with a peak at 6100A. It was, therefore, of interest to study the question further.

The phosphors were prepared by firing mixtures of C.P. ZnO, BeO, $MnCO_3$, and $SiO_2 \cdot xH_2O$, in the proportions shown in Table I, at approximately 1200° until their x-ray diffraction patterns showed the same lattice spacings found by Fonda[3] in his fusion preparations. This was taken to indicate substantially complete reaction of the components. The spectral energy measurements were made on conventional fluorescent lamps internally coated with the above preparations. Phosphor excitation in such lamps is primarily caused by the 2537A resonance radiation generated in the low pressure mercury discharge.

The determinations were based on the comparison of the fluorescent lamp output with that of an incandescent source of known spectral distribution. The experimental arrangement consisted of a glass double monochromator equipped with a constant speed motor drive and a revolution counter. By calibration with a high pressure mercury lamp, counter readings were converted to wave-lengths. The receiver was a Cs-Ag-O vacuum photo-tube. Its response was amplified by a Roberts type d.c. amplifier[4] connected to an Esterline-Angus recording milliammeter. A continuous record of photo-tube response *vs.* counter reading (wave-length) was obtained in this way, the short period of the recording instrument allowing the range from 4000–7000A to be scanned in seven minutes. The fluorescent lamps were mounted close to the entrance slit without interposition of a focusing lens. To approximate these conditions with the incandescent standard, a block of magnesium carbonate, placed close to the entrance slit and illuminated by the incandescent lamp, served as a secondary source. The standard source was a low voltage, ribbon-filament tungsten lamp operated at a temperature of 2900°K. Slit widths were kept fixed throughout the runs, adjustment of the amplifier gain being employed to obtain recorder deflections of proper magnitude. This adjustment could be made without interrupting the runs. The slit width used corresponded to a $\Delta\lambda$ of 8 mμ at 4000A increasing to 50 mμ at 7000A. Results obtained with this arrangement have been found to agree very closely with data obtained from point by point measurements at a constant $\Delta\lambda$ of 10 mμ. Results of the determinations are shown in Fig. 1, E_{5600} being set equal to 100 for all curves.

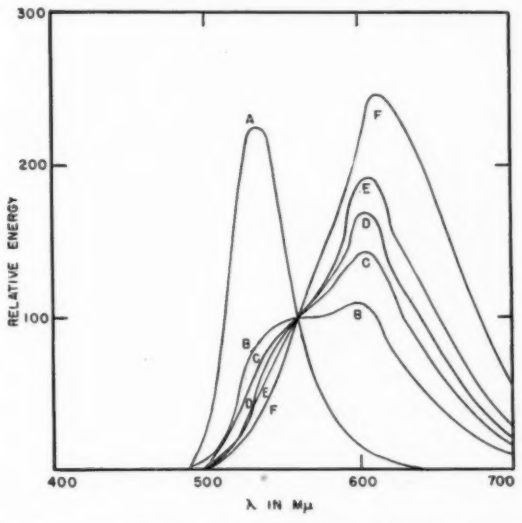


FIG. 1. Spectral energy distribution of luminescence of samples in Table I.

[3] G. R. Fonda, J. Phys. Chem. **45**, 282 (1941).

[4] S. Roberts, Rev. Sci. Inst. **10**, 181 (1939).

Approximate quantum efficiencies of the various phosphors are given in Table II. In the calculation of these values it is assumed that excitation is entirely caused by 2537A radiation, and that the efficiency of conversion of electrical energy to this radiation is 60 percent in a 40-watt, 1.5-inch diameter, fluorescent lamp. Measurement of the luminous efficiencies of such lamps containing the zinc orthosilicate and zinc beryllium silicate phosphors were made using an integrating sphere, the luminous efficiency of the mercury discharge, about 5 lumens per watt, being subtracted from the measured values to get the efficiencies of the phosphors themselves. Calculations of the quantum efficiencies were made by the method of Thayer,[5] using his data on the absorption of 2537A and luminescent radiation by the phosphor coating, visible light absorption by the lamp bulb, and light loss by absorption in the lamp ends. It should be noted that milling of the phosphors is necessary in preparing suspensions for coating the lamps. Since the grinding of phosphors destroys some of their luminescent centers, the efficiencies of the above preparations were originally higher than the values given.

It may be seen from the figure that the introduction of 10 mole percent of beryllium into zinc orthosilicate suppresses the characteristic willemite luminescence and creates a new emission band with a peak at about 6100A. The quenching of the 5250A band and the growth of the 6100A band are both fostered by increase of the manganese concentration at this Zn : Be ratio, and by further increase of the beryllium to 20 mole percent. A high quantum efficiency is maintained through all these changes.

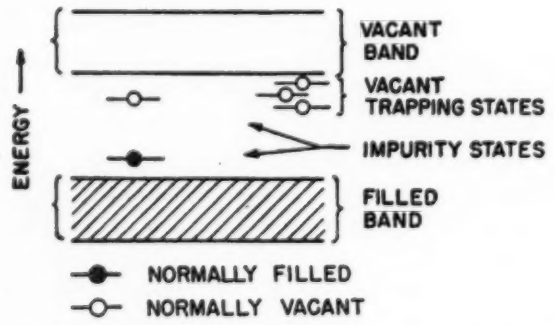


FIG. 2. Energy level diagram for silicate phosphors according to R. P. Johnson.

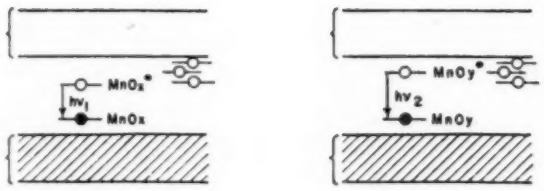


FIG. 3.

DISCUSSION

The widespread occurrence of manganese as an activator and its association with green, yellow, or red emission bands in a wide variety of host structures suggest that the transition responsible for luminescence takes place within the manganese atom or ion itself or within a coordination group consisting of the manganese ion and its immediate neighboring cluster of anions.[2,6] This viewpoint is likewise supported by studies of the luminescence decay and of the photo-conductivity of manganese-activated phosphors.[7,8] From such studies Randall and Wilkins[8] have concluded that the initial exponential decay of all these phosphors is caused by capture of an electron in an excited state of the Mn^{++}-ion, and is characterized by the lifetime of this excited state. This is essentially the scheme proposed in Johnson's[7] model of silicate* phosphors, illustrated in Fig. 2, if the normally-empty upper impurity level of this diagram is interpreted as the excited Mn^{++} state. The principal features of Johnson's model are the assumptions that each activator creates *two* localized states in the forbidden region of the host crystal, the lower level being normally filled and the upper level normally vacant, and that the latter is accessible to an electron only if the associated lower level is also vacant at the same time. The transition from the upper to the lower of these two impurity states, resulting in fluorescence, is assumed to be

[5] R. N. Thayer, Electrochem. Soc. Preprint **87**, 25 (1945).

[6] S. H. Linwood and W. A. Weyl, J. Opt. Soc. Am. **32**, 443 (1942).

[7] R. P. Johnson, J. Opt. Soc. Am. **29**, 283 and 387 (1939).

[8] J. T. Randall and M. F. Wilkins, Proc. Roy. Soc. **184**, 347 (1945).

* Since the published data forming the basis of Johnson's picture deal entirely with manganese-activated silicate phosphors, his model is here considered to represent manganese-activated phosphors in general rather than silicate phosphors in general.

"radioactive in character, little affected by external parameters, and requiring an average time of the order of 10 milliseconds." The model leads to the observed two-stage decay of manganese activated silicate phosphors, accounting for the initial exponential decay and also, *via* the trapping states, for the longer-lived second stage which resembles the decay of sulfide phosphors.

We may incorporate in Johnson's scheme considerations of the luminescence color by assuming further that the average energy difference between the two localized states associated with the manganese activator is a function principally of the number and kind of anions in the immediate coordination sphere of the manganese (cf. Fig. 3), the influence of the host lattice on this separation being of a much smaller magnitude. The manganese-activated phosphors as a class are accordingly considered to be intermediate in type between the (Zn, Cd)S phosphors, where the emitted frequencies are determined by the position of the activator level with respect to the conduction band of the host structure, and the rare earth, Cr^{+++}, and UO_2^{++}-activated phosphors, whose luminescence emission is very markedly characteristic of the activators themselves.

The influence of the coordination number of manganese on the luminescence colors of Mn-activated materials has been considered by Linwood and Weyl.[6] From a study of the fluorescent silicate, borate, and phosphate glasses they conclude that the "green" fluorescence arises from a fourfold coordination of Mn^{++} with oxygen, while the "red" fluorescence arises from sixfold coordinated manganese. In the first case the manganese is a "glass network former," i.e., it occupies a position in the glass structure similar to that of silicon; in the latter case it is a "glass network modifier," occupying interstitial positions throughout the glass structure, such as alkali or alkaline earth ions do. Corresponding to the first-named role in the glass structure, the bonding of manganese to oxygen is strong, symmetrical, and not greatly affected by thermal vibrations. In interstitial positions, corresponding to the "network modifier" role, the Mn−O bonding is weaker, less symmetrical, and more easily disturbed by thermal vibrations. The effects of temperature and concentration of activator on the intensities of the green and red fluorescence bands are in qualitative agreement with this viewpoint.

In the extension of their picture to crystalline phosphors, Linwood and Weyl point out that manganese ions situated at crystal defects would have an environment similar to that found at "network modifying" positions in glasses. They reach the conclusion that in general "the long wave emission of Mn^{++} occurs whenever the ion is brought into positions of higher kinetic energy, either interstitial holes or flaws in the crystal structure."

The application of their coordination scheme to the willemite derivative phosphors was attempted by the above authors with partial success. The green luminescence of Zn_2SiO_4:Mn itself was readily and explicitly accounted for by the direct substitution of Mn^{++} for Zn^{++}. Since the Zn^{++} ions in willemite* are surrounded by four oxygens,[9] this substitution gives rise to the MnO_4 groups responsible for the green emission. The appearance of the red component of the luminescence in $(Zn, Be)_2SiO_4$:Mn was, however, explained in only general terms, being presumably caused by distortion of the willemite structure by replacement of Zn^{++} ($r=0.83$A) with the smaller Be^{++} ($r=0.34$A), or to changes in the symmetry of the Mn^{++} environment or to lattice defects produced by this substitution. A more specific picture of the possible origin of the "red" centers in the zinc beryllium silicate phosphor is offered in the following discussion.

Over rather wide limits of manganese concentration, the spectral energy distribution of the Zn_2SiO_4:Mn luminescence changes but little from the distribution shown in Fig. 1, curve *A*. The optimum concentration of manganese in this phosphor is about one percent by weight (0.04 mole Mn per mole Zn_2SiO_4). Further increase in activator concentration results in an efficiency loss, together with a broadening of the emission band and a slight shift of its peak toward the red. There is, however, no indication of the growth of a new band at longer wave-lengths.

* The same is true for the Be^{++} ions in phenacite, Be_2SiO_4, which is iso-structural with willemite. Be_2SiO_4 and Zn_2SiO_4 form limited solid solutions containing up to 30 mole percent Be_2SiO_4 (references 8 and 3).

[9] W. L. Bragg and W. H. Zachariasen, Zeits. f. Krist. **72**, 518 (1930).

With a very high manganese concentration, corresponding to complete quenching of the room temperature luminescence, Kröger[2] was able to observe a bright luminescence at $-180°$. This low temperature luminescence consisted of both the usual green band and a feebler band peaking at 6100A, apparently identical with the emission brought out at room temperature by incorporation of beryllium into the structure.*

According to the strictest interpretation of the coordination picture, the 6100A band requires the existence of interstitial manganese with a coordination number greater than four. The appearance of this band in zinc orthosilicate would therefore indicate the presence of such interstitial manganese even in this unsubstituted structure. The preference of manganese for the role of "network former" (fourfold coordination with oxygen) in this structure is shown, however, by the high activator concentration required before the appearance of the 6100A emission, this concentration being far above the optimum for the green luminescence. The interstitial manganese might be produced by the occasional substitution of Zn^{+2} for Si^{+4} in the willemite structure, resulting in the formula

$$\left.\begin{matrix} Zn^{++}{}_{(2-a)} \\ Mn^{++}{}_{(x+a)} \end{matrix}\right| \left.\begin{matrix} Si^{+4}{}_{(1-x)} \\ Zn^{++}{}_{(x)} \end{matrix}\right| O_4, \qquad (1)$$

where $x Zn^{++}$ replace $x Si^{+4}$, $x Mn^{++}$ are in interstitial positions as a consequence, and $(2-a)Zn^{++}$ and $a Mn^{++}$ are in the normal fourfold coordinated cation positions. Because of the great difference in size and polarization properties between Zn^{++} ($r=0.83$A) and Si^{+4} ($r=0.39$A), the extent of this substitution would be small and the green emission band would predominate.*

This type of substitution should be considerably enhanced in the zinc beryllium silicate phosphors, owing to the great similarity in size and polarization properties of Be^{++} ($r=0.34$A) and Si^{+4}. In this case a possible structural scheme might be the following:

$$\left.\begin{matrix} Zn^{++}{}_{(2-a-b)} \\ Be^{++}{}_{(a)} \\ Mn^{++}{}_{(x+b)} \end{matrix}\right| \left.\begin{matrix} Si^{+4}{}_{(1-x)} \\ Be^{++}{}_{(x)} \end{matrix}\right| O_4, \qquad (2)$$

where $x Be^{++}$ replace $x Si^{+4}$; $a Be^{++}$, $b Mn^{++}$, and $(2-a-b)Zn^{++}$ are in the normal cation positions; and $x Mn^{++}$ are in interstitial positions demanded by the $Be^{++} \rightarrow Si^{+4}$ substitution. The Zn^{++} (and Be^{++}) may, of course, compete with the Mn^{++} for these interstitial positions, especially when the Mn^{++} concentration is low. As this is increased, however, the Mn^{++} ions have a correspondingly greater probability of occupying the interstitial holes, and being larger than the Zn^{++} and Be^{++}, may even be preferentially housed in these positions. This would account for the increased 6100A emission with increase in manganese concentration at a constant Zn:Be ratio. To account for the rapid decrease of the "green" band intensity by introduction of beryllium, it may be necessary to assume also that excitation of the "interstitial-Mn^{++}" centers is more probable than excitation of the "substitutional-Mn^{++}" centers.

Various other possibilities are opened up by considering the effect of different substituents for silicon in the willemite structure. According to the foregoing considerations, replacement of Si^{+4} by a suitable tetravalent element, such as germanium, should have no great effect on the spectral energy distribution of the resulting phosphor, for no interstitial manganese ions are made necessary by this substitution. According to the data of Leverenz[1,10] this is apparently true, the luminescence of Zn_2GeO_4:Mn consisting of a

* The red emission made observable by cooling should not be interpreted as the development of a new type of emitting center resulting from structural changes induced by the low temperature. Except for the lattice contraction due to the temperature decrease, we shall assume that the structures of both the red-emitting and green-emitting centers are the same as they would be at room temperature. With the high activator concentrations employed, the thermal vibrations corresponding to room temperature result in a quenching of the luminescence, both red and green, at this temperature. The function of the cooling is merely to reduce the energy exchange between both types of center and the host lattice, in this way making both fluorescence processes possible. It should be mentioned that although cooling and substitution of beryllium for zinc both operate to contract the structure, this cannot be the reason for the change in luminescence spectrum. For if the spectral change were merely a consequence of the lattice contraction, one would expect a gradual shift of the emission peak with decreasing temperature or with increasing beryllium concentration (cf. the (Zn, Cd)S phosphors) instead of the appearance of a new band.

* An identical substitution was postulated by Linwood and Weyl to account for the formation of yellow-fluorescent zinc silicate, a polymorphic form of willemite, with presumably a cristobalite-like structure.

[10] H. W. Leverenz, RCA Rev. **5**, 131 (1940).

single band like that of Zn_2SiO_4:Mn, with a small shift of the peak to 5370A.* The same should hold true for any other tetravalent substituent for Si^{+4}, provided that: (a) the willemite structure is retained in the solid solution, (b) the substituent does not act as a "poison" for the luminescence, and (c) the substituent does not introduce a luminescence band of its own.

The substitution for silicon of elements having a higher valence may likewise be approached from this standpoint. For example, Froelich[11] has found that incorporation of arsenic or arsenate in willemite causes an "exaggeration" of its phosphorescence. Froelich and Fonda[12] have shown that the action of arsenic consists in increasing the number of trapping states. Substitution as As^{+5} ($r=0.4$A) for Si^{+4} in fourfold coordination with oxygen should be quite possible, giving either

$$\left.\begin{matrix} Zn^{++}_{(2-a-(x/2))} \\ Mn^{++}_{(a)} \end{matrix}\right| \left.\begin{matrix} Si^{+4}_{(1-x)} \\ As^{+5}_{(x)} \end{matrix}\right| O_4 \qquad (3)$$

or

$$\left.\begin{matrix} Zn^{++}_{(2-a)} \\ Mn^{++}_{(a)} \end{matrix}\right| \left.\begin{matrix} Si^{+4}_{(1-(5x/4))} \\ As^{+5}_{(x)} \end{matrix}\right| O_4. \qquad (4)$$

The occasional AsO_4 groups distributed through the structure would thus introduce into the crystal positive holes where electrons could be trapped. In (3) a Zn^{++} ion vacancy should occur for every two As^{+5} atoms substituting for Si^{+4}; in (4) a Si^{+4} vacancy should occur for every four substituting As^{+5} atoms, the positive holes being attracted to the vacant lattice point.

Similar considerations should apply to the substitution of various elements for zinc in the willemite structure. Interstitial cation positions would be demanded by replacement of zinc by a suitable monovalent ion, such as Li^{+1} ($r=0.78$A)* or Na^{+1} ($r=0.98$A), resulting in the following structural scheme.

$$\left.\begin{matrix} Zn^{++}_{(2-a-x)} \\ Li^{+}_{(x)} \\ Mn^{++}_{(a+x/2)} \end{matrix}\right| Si^{+4} \left| O_4. \right. \qquad (5)$$

In (5), xLi$^+$ and aMn^{++} substitute for $(a+x)$Zn^{++} in the normal cation positions, and $x/2$Mn^{++} are in interstitial positions. Here again it is possible for the Zn^{++} to compete with Mn^{++} for the interstitial positions, and it may be that the red emission will be observed only at high manganese concentrations. The manganese concentrations required for the appearance of the red band in this case should be lower than that needed in the absence of lithium substitution, however. Trivalent or quadrivalent substituents for Zn^{++} should lead to an increase in phosphorescence by the same method that the substitution of As^{+5} for Si^{+4} enhances this property.

It is hardly necessary to mention that the "red" centers can arise in phosphors, without the rather complicated scheme described above for the willemites, by direct substitution of Mn^{++} for the cation of structures where the latter is normally sixfold coordinated with oxygen. The red luminescence of Mg_2SiO_4:Mn is an example of such a case.

The foregoing considerations may apply also to halide, sulfide, and other types of manganese-activated phosphors, where the appropriate values of n in the MnX_n clusters would have to be determined.

SUMMARY

The effectiveness of manganese as an activator in a large number of different host structures, and the repeated occurrence of either red or green luminescence bands, or both, in such phosphors indicate that the electronic transition giving rise to luminescence takes place within the coordination group of manganese and its immediate anion cluster. This view is supported by the exponential decay characteristics of these phosphors, and by the apparent absence of a connection between the luminescence and the photo-conductivity of these materials. A tentative model of manganese activated phosphors is obtained by combining Johnson's energy level scheme for the decay of silicate phosphors with Linwood and Weyl's results on the effect of

* Variation of manganese concentration in Zn_2SiO_4 itself causes a shift in luminescence peak from 5230 to 5315A (reference 1). The shift in peak between Zn_2GeO_4 and Zn_2SiO_4 is thus a minor effect, probably connected with expansion of the lattice by the $Ge^{+4} \rightarrow Si^{+4}$ substitution.

[11] H. C. Froelich, U. S. Patent 2,206,280 July 2, 1940.

[12] H. C. Froelich and G. R. Fonda, J. Phys. Chem. **46**, 878 (1942).

* The fourfold coordination of Li^+ with oxygen occurs in Li_2WO_4 and Li_2MoO_4, both of which are "model compounds" of Zn_2SiO_4 (reference 9).

coordination number on the luminescence spectrum of manganese-bearing crystals and glasses. Spectral energy distribution measurements of the luminescence of Zn_2SiO_4:Mn and of various samples of $(Zn, Be)_2SiO_4$:Mn are reported and discussed from the above viewpoint.

ACKNOWLEDGMENT

The writer is greatly indebted to Professor A. R. von Hippel, Laboratory for Insulation Research, M. I. T., for his kind permission to use the monochromator and associated equipment employed in the measurements.

Recent Ozone Investigations

ARTHUR W. EWELL*

Worcester Polytechnic Institute, Worcester, Massachusetts

(Received February 16, 1946)

This paper is supplementary to one which appeared in the December, 1942 issue of this journal. In many applications of bactericidal lamps in air conditioning—such as food preservation, processing, etc.—low concentrations of ozone are desirable to protect surfaces and air pockets not reached by the radiation. A few tenths of one part of ozone per million parts of air is generally sufficient and is obtainable by selecting a lamp with a wall which transmits in addition to the very bactericidal 2537A radiation a few percent of 1850A. The latter is absorbed by the oxygen in the air forming ozone. The reciprocity law (amount of bactericidal action proportional to concentration×duration of ozone) is applicable for moderate concentration ranges. The bactericidal effect of ozone is nearly independent of humidity for surface infection of foods, etc., while high humidity is required for destruction of air borne bacteria. No oxides of nitrogen are produced by high transmission glass or quartz lamps. The ozone output of a lamp is reduced by decomposition by intense 2537 radiation either from the same lamp or from another lamp nearby. Moderate increases in the temperature of a high transmission lamp does not affect the ozone output but increases the dissociation constant until, at 150°F, practically all the ozone is immediately decomposed. The ozone concentration produced by a lamp is substantially reduced by rise in relative humidity, which moderately absorbs 1850 radiation and decomposes much of the ozone produced. Approximate reflection coefficients for 1850 radiation at various surfaces have been obtained by an indirect method.

THIS paper supplements "Production, Concentration and Decomposition of Ozone by Ultraviolet Lamps" which appeared in the December, 1942 issue.[1]

OZONE IN AIR CONDITIONING

The chief purpose of the ozone produced by a bactericidal lamp is to supplement the much more germicidal 2537A radiation at infected surfaces not reached by direct, reflected, or scattered radiation, and air-borne infection in stagnant air pockets. Ozone is aided in reaching such pockets by a high diffusion coefficient and in general spreads throughout an enclosure faster than gravitational air currents. On account of the bactericidal contribution of the accompanying 2537 radiation the required ozone concentrations are much lower than required when the ozone is produced by ozonators.

The ozone produced from the oxygen in the air by radiation of wave-length 1850A, emitted by high transmission ultraviolet lamps, is more effective than the accompanying 2537 radiation in destroying both odor and taste molecules in vapors, and ethylene gas, the latter accelerating ripening of some fruits. The ozone also increases the aroma of small fruits such as strawberries.[2]

If "house-keeping" is satisfactory, very low concentrations—a few tenths of one part per million (p.p.m.)—are, in regions shaded from 2537 radiation, sufficient to inhibit the growth of surface infection, and air-borne infection—the latter only if the relative humidity is above 40 percent—and to destroy in vapors from foods the molecules which contribute odors and tastes and cause cross-contamination and an objectionable general odor. In the "Tenderay" process for tenderizing meat, 0.2 p.p.m. suffices for control of infection on shaded surfaces; 0.3 p.p.m. is sufficient in cheese aging rooms. Domestic re-

* Consultant, Westinghouse Electric Corporation.

[1] A. W. Ewell, J. App. Phys. **13**, 759–767 (1942).

[2] A. W. Ewell, *Refrig. Data Book* (1946 Edition), pp. 193–199; Refrig. Eng. **48**, 381–383 (1944).

frigerators require not over 0.1 p.p.m. from special refrigerator Sterilamps.[3] The 1850 radiation required to produce such concentrations is only a few percent of the total 2537 and 1850 radiation.

Although high concentrations of ozone are objectionable in general air conditioning, low concentrations accompanying 2537 radiation contribute to air purity. Yaglou[4] found that a concentration of 0.015 p.p.m. in a room crowded with people of "the poorest economic class" reduced body odor to such an extent that to remove the odor the required air changes with outside air could be reduced 50 percent.

The all important factor in applications of ozone is the ozone concentration, c (grams of ozone per cubic meter), determined by the rate of ozone production, per unit volume of inclosure, m (g of ozone produced per minute per m^3), and the decomposition coefficient, k ($k=0.7/T$ where T is the time in minutes, after the ozone is shut off, for the concentration to drop to one-half).

$dc/dt=m-kc$.* At equilibrium $dc/dt=0$, and $m=kC$, where C is the equilibrium concentration.

COMPARATIVE BACTERICIDAL EFFECT OF OZONE AND 2537 RADIATION

The reciprocity law, which requires that the percent killing be proportional to the concentration×duration of ozone (p.p.m.×hr.) holds for limited concentrations. For 50 percent killing of *E coli* on a surface, the writer has found that the product, p.p.m.×hr., is about 1.8 for ozone concentrations between 0.5 and 3 p.p.m.[5] Other investigators have found results of the same order of magnitude for mould spores.[6] The amount of 2537 radiation to produce 50 percent killing of *B coli* is about 5 clicks of a Rentschler Tantalum Click meter.[7] Therefore, for ozone concentrations of the order of 1 p.p.m., 1.8 (p.p.m.×hr.) and 5 Tantalum cell clicks have approximately equal bactericidal effects or the bactericidal effect upon surface infection of 1 p.p.m. of ozone for 1 hr. is approximately equivalent to 3 Tantalum cell clicks of 2537 radiation.

TABLE I.

Relative humidity	Ozone concentration p.p.m.	% kill
28%	4.3	0
36	3.5	0
46	4.	34
52	4.	41
70	3.5	40
92	0.7	83
95	3.5	99

BACTERICIDAL EFFECT AND HUMIDITY

Rentschler and Nagy[8] found that the bactericidal effect of ultraviolet radiation is independent of the surrounding humidity for both surface and air-borne infection. There is considerable evidence that ozone is only germicidal in the presence of water.[9] If bacteria, mould, etc., are on surfaces, which are not dried out, of meat, cheese, fruit, etc., or on a nutrient in a Petri plate, there will always be high humidity for a short distance from the surface. The surface infection is, therefore, associated with sufficient water so that the percentage killing by ozone is quite independent of the relative humidity of the air.

AIR-BORNE INFECTION

Wm. J. Elford and Joan van der Ende[10] found, using very low ozone concentrations and long exposures, that at relative humidities below 45 percent, the germicidal power of ozone upon air-borne infection was negligible, while at high humidities it was very great even for concentrations far below 0.1 p.p.m., if the droplets were not so large as to prevent access to the bacteria, and did not contain excessive absorbing organic matter.

Such dependence upon humidity of the bactericidal effect of ozone upon an air-borne infection is so important in many germicidal applications that it seemed desirable to check these (the only published results known to the writer) by entirely different experimental procedure, with higher ozone concentrations and shorter times.

[3] A. W. Ewell, Refrig. Eng. **50**, 523, 524; 560, 561 (1945).
[4] Yaglou, Heating, Piping and Air Conditioning **11**, 648–653 (1939).
* Beyond a few inches from the lamp—see Table II.
[5] A. W. Ewell, Refrig. Eng. **41**, 331–334 (1941); Refrig. Eng. **42**, 283–286 (1942); Refrig. Eng. 159–161 (1943).
[6] Smock and Watson, Refrig. Eng. **42**, 97–101 (1941).
[7] Rentschler, J. Am. Inst. Elec. Eng. **49**, 113–115 (1930).
[8] Rentschler and Nagy, J. Bact. **44**, 85–94 (1942).
[9] Kaess, Chem. Zeits. **62**, 365 (1938).
[10] W. J. Elford and J. van der Ende, J. Hygiene, London **42**, 240–265 (1942).

TABLE II.

Distance between axes of lamps	Reduction in ozone
1½ in.	23%
1¾	13%
2½	4%
3	0%

An enclosure of about 1.4 m³ was employed. The ozone was produced by a 1000-cycle Siemens & Halske ozonator, the rate of production being controlled by the primary voltage. When a quite constant and uniform concentration was attained, 10–20 cc of a broth culture was sprayed into the top of the booth. The air-borne infected droplets fell slowly. Ten minutes after spraying, sterile culture plates were exposed at the bottom of the box by withdrawing a slide for 1 minute, collecting bacteria which had been sinking through the air and ozone 10–11 minutes. Controls were secured by similar procedure with no ozone. Killing of bacteria during this minute after settling on the plate was determined by exposing plates with known seeding to the same ozone concentration for one minute. The percentage thus killed was subtracted from that of the settling bacteria. The results are presented in Table I. The temperature throughout was close to 70°F.

The manipulation in both this and in the English method was difficult, resulting in some variation in results with both. The agreement is, however, sufficient to justify the conclusion that the *killing of air-borne infection by ozone* is nil at low humidities, moderate at 50 percent relative humidity, and very high when the relative humidity approaches 100 percent. (Note the high killing by only 0.7 p.p.m. at a high humidity.)

DECOMPOSITION OF OZONE BY 2537 RADIATION

Two similar lamps, both producing the same amount of 2537 radiation, but one giving a large ozone output and the other none, were placed side by side in still air. Table II shows the reduction in ozone from the first lamp by additional intense 2537 radiation near the second lamp. With both lamp and silent discharge sources of ozone the output increases when an air current is applied which reduces the length of time that the ozone is subject to the deozonizing as well as the ozonizing influence near the source. With brush discharge ozonators the output increases to a constant value at a moderate air speed. With a high transmission lamp as a source of ozone, the yield of ozone increases to a maximum as the transverse air velocity is increased, followed by decrease due to cooling of the lamp. With quartz Sterilamps this maximum occurs at about 100 feet per minute transverse air velocity.

With *increase in humidity* the ozone output from a lamp decreases moderately because of absorption of the 1850 radiation by water vapor. The ozone concentration is further decreased because of increase in the decomposition coefficient.

An *increase in temperature* from 6°C to 31°C (88°F) decreased the ozone produced by lamps to about one-half.[1] Since the temperature in rooms, and particularly ducts, where lamps are installed may be much higher, the following investigations were made:

A high transmission, No. 793, Sterilamp—in an enclosure—was heated in an electric furnace. Tests which were only qualitative indicated a decrease in ozone concentration to zero as the temperature was raised from 70° to 170°F. The current through the lamp only varied between 38.5 and 40 milliamperes. By opening one side of the enclosure the ozone producing 1850 radiation was measured with a Rentschler platinum click meter. Throughout the above range in temperature the 1850 radiation remained constant within 10 percent. (At higher temperatures and mercury vapor pressures, the 1850 output falls off.) Since the ozone output has always been found proportional to the platinum cell readings, it was evident that the decrease in ozone concentrations was caused by increased decomposition. Such was proved to be the case by sending the same initial ozone concentration through an enclosure at different

TABLE III.

Temperature (F)	Ozone concentration
57°	3.
77	2.7
88	2.2
97	1.7
108	1.3
123	0.7
150	0

Above 100° the accuracy was low.

temperatures, then cooling the ozone-air mixture and determining the concentration. (See Table III.)

OXIDES OF NITROGEN

The writer has never found with tetrabase paper any evidence of oxides of nitrogen in the vicinity of lamps producing ozone[11] and none has been found in more careful investigations by others. The most comprehensive search for such oxides, using the most sensitive reagents, was conducted by Kohn-Abrest who found that the only gaseous product produced by glass or quartz lamps was ozone.[12]

REFLECTION OF 1850 RADIATION

Reflection coefficients of surfaces for 2537 radiation, at various angles of incidence, can be determined with a Rentschler Tantalum Click Meter. The intensity of reflected 1850 radiation from a bactericidal lamp at moderate angles of incidence is, however, too low for accurate measurements. Knowledge of reflection coefficients for different surfaces is important for choice of reflection materials for fixtures of ozone producing lamps. The following indirect method gave results of qualitative value:

A 9-mm quartz, 4-in. lamp (Fig. 1) was mounted horizontally near the center of a 1.4 m³ enclosure, between horizontal glass plates 6 in. ×8 in. A fan at the lamp level gave an air velocity of 100 F.P.M. The equilibrium ozone concentration was determined by test papers*

TABLE IV.

Black photographic paper	23%
Aquadag coated glass	29%
Bare glass	34%
Alzak	57%

TABLE V.

Distance between plates	Approximate unobstructed volume of air	Average ozone concentration p.p.m.	Ozone concentration compared with no obstruction
No plates	100%	2.0	100%
8 in.	50	1.7	87%
6 in.	40	1.6	80%
4 in.	30	1.4	70%
2 in.	15	1.1	55%

[11] A. W. Ewell, News. Ed. Am. Chem. Soc. **19**, 1102 (1941).

[12] Kohn and Abrest, Comptes rendus **184**, 482–484 (1927).

* It is the writer's experience that the most feasible method for measuring concentrations, 0.4–20 p.p.m., is exposure of potassium iodide-starch papers, of proper texture, the ozone concentration being determined from the time, after wetting, required to match one of three standard colors. Papers with colors are calibrated by careful determination of the ozone concentration by titration. Absorption of 2537 radiation as used by Koller [J. App. Phys. **16**, 816–820 (1945)] is preferable for high concentrations. Until more sensitive tantalum cells are available, this method is inaccurate for concentrations below 2 p.p.m. For this concentration, at least 200 cm of air and ozone are necessary.

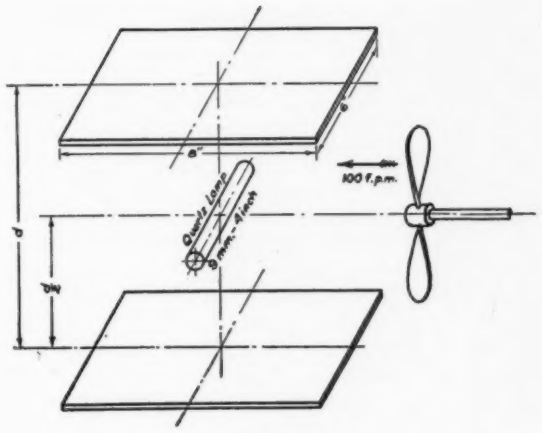


FIG. 1.

suspended from sounds at various locations in the enclosure.

The 1850A ozone producing radiation, E, is absorbed by the oxygen of the air (producing ozone) according to the equation[1]

$$E/E_0 = e^{-.054d},$$

where E is d cm farther from the lamp than E_0. For a distance of 4 in. $1-E/E_0$ is about 0.4. When the plates were 4 in. above and below (d in Fig. = 8 in.), the unobstructed 1850 radiation was approximately 50 percent. The total ozone producing radiation compared with no plates was, therefore,

$$0.5+0.4\times0.5+0.6\times0.5R=0.7+0.3R=C/C_0,$$

when R is the fraction of the incident radiation reflected at angles of incidence, approximately 0–45°, C the ozone concentration in the inclosure and C_0 the concentration without any plates.

Table IV presents the values of R for various surfaces given by this approximate method.

Varying the distance between the alzak plates, which were equal distances from the lamp above and below, gave the results in Table V.

The high values for small unobstructed volumes of air are caused by very great reflection at high angles of incidence.

Properties of Welded Contact Germanium Rectifiers

H. Q. North
General Electric Research Laboratory, Schenectady, New York
(Received May 21, 1946)

In the contact rectifier described, the cat whisker has been solidly welded to a germanium surface. The welding process has brought about several extraordinary properties not found in the usual point rectifier: (1) Mechanical stability is achieved without the use of a compound to fill the cartridge. Protection against moisture is gained through the use of a glass-metal seal. (2) The forward dynamic resistance of the microwave unit at 0.5 volt is exceptionally low (2 to 4 ohms), and the back resistance at 0.5 volt is high (0.01 to 1.0 megohm). These values of resistance are accompanied by a low contact capacity at zero bias (below 0.2 $\mu\mu$f). (3) The forward d.c. characteristic follows the law, $I_F = I_0 \exp(\alpha V_F)$ over more than three decades of current. (4) The low forward resistance is responsible in part for conversion *gain* when the crystals are mismatched to microwave radiation at the r-f terminals. (5) Specially welded crystals used as microwave harmonic generators have yielded more than twenty times the harmonic (4-mm wave-length) power output of standard silicon crystals.

1. INTRODUCTION

DURING recent years the need for high sensitivity, low loss detectors at microwave frequencies has reopened the development of contact rectifiers of the point-plane type. The unit described in this paper resulted from an extended investigation carried out in the Research Laboratory of the General Electric Company.[1] The overall purpose of the work was to investigate germanium as a possible improvement over silicon for radar mixers. Only that part of the project which has yielded a new crystal of unusual properties is treated here. An attempt has been made in this presentation to include only enough background of the subject to make the results understandable.

The most important result of our germanium work has been the development of a crystal rectifier in which the cat whisker is firmly welded to a polished germanium surface. The rectifier assembly has been hermetically sealed in a microwave cartridge by means of a glass-metal seal.

The welded microwave rectifier when used as a mixer at 3-cm wave-length has a noise figure as low as or lower than that of the best silicon mixer crystal. It has in addition some extraordinary properties. In the first place, when an impedance mismatch exists between signal source and crystal with local oscillator applied, *amplification* may actually occur. The increase in signal is accompanied by an increase in noise, however, and the value of the crystal as an amplifier is questionable. Secondly, in the forward direction the unit follows closely the law $I_F = I_0 \exp(\alpha V_F)$ over three decades of current. Here I_F and V_F are the forward current and voltage, respectively. I_0 and α are constants. If the exponent αV_F is changed to $\alpha(V_F - I_F R_s)$, where R_s is a constant (the spreading resistance), the plot of $\log I_F$ *vs.* $(V_F - I_F R_s)$ is followed closely over more than 5 decades of current. Finally, specially welded germanium crystals have proved superior to commercially obtainable silicon units as generators of both the second and third harmonic of 1.2-cm radiation.

2. THE CRYSTAL CARTRIDGE ASSEMBLY

The cartridge in which all of the microwave tests were made is shown in Fig. 1. Its external dimensions have been made to conform to Army-Navy specifications for the 1N26 type of crystal rectifier. A 0.0015″ diameter platinum-10 percent ruthenium wire is first sharpened to a point of less than 0.00002″ radius. It is then kinked and spot-welded as shown to the silver-plated steel

[1] This work was done in part under Contract No. OEMsr-1377 between the General Electric Company and the Office of Scientific Research and Development and under subcontract with the Massachusetts Institute of Technology, Prime contract OEMsr-262. The OSRD assumes no responsibility for the statements contained herein.

This paper is a condensation of material presented by the author to the University of California at Los Angeles in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Physics.

center conductor. The assembly is then sealed to the outer shell with a low loss glass which matches the steel in expansion.

The crystal is a small pellet of treated germanium, soldered to the plug as shown, and polished slowly, first with 600-mesh alundum and finally with magnesium oxide. Electron diffraction photographs indicate that if this slow polishing is done under water, little or no debris remains on the surface.

The germanium used in the welded-contact microwave unit is pure germanium to which 0.2 atomic percent of antimony is added. The addition of this "donor" changes the bulk resistivity from approximately 10 ohm-cm to about 0.005 ohm-cm. The theory of conduction accounting for this 2000-fold change in resistivity has been treated elsewhere[2,3] and need not be reviewed here. It should be mentioned, however, that rigid control of impurities is vital to the construction of good rectifiers. According to theory, approximately 1 atom of antimony in 2000 of germanium is active in producing the 2000-fold change of resistivity mentioned. The remainder of the antimony added is deposited at the grain boundaries.

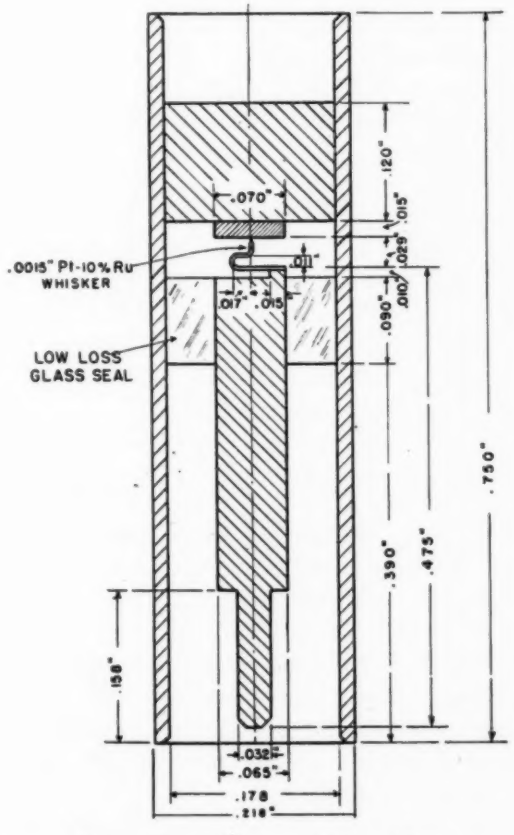


FIG. 1. Germanium crystal cartridge.

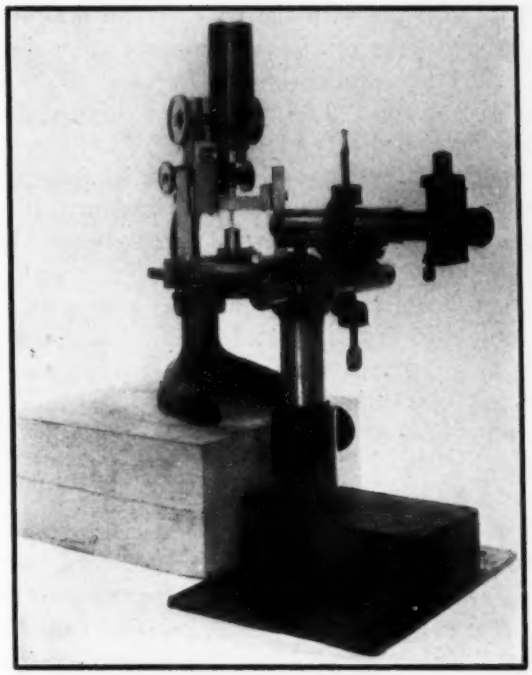


FIG. 2. Assembly to determine force on the contact as a function of whisker deflection.

3. THE WELD AND ITS PHYSICAL PROPERTIES

Early in our research on germanium rectifiers it was found that conversion loss and noise could be greatly decreased by subjecting the units to forward direct currents as high as 100 milliamperes. Contact areas were large at this stage of development, however, and no noticeable welding occurred. Later in the work, condenser discharges were tried, and actual welds were photographed. Because of the large contact areas existing at that time, the condenser-welded units showed poor microwave performance. After the adoption of the present cartridge, a series of careful measurements indicated that contact force should initially be as low as possible and that welding currents should be high for best performance. Best results were obtained when the plug to which the crystal was soldered was pressed in 0.0005" after contact

[2] N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Clarendon Press, Oxford, 1940).
[3] F. Seitz, *The Modern Theory of Solids* (McGraw-Hill Book Company, Inc., New York, 1940).

had been made with the whisker. An optimum welding current of 250 milliamperes was passed through the contact in the forward direction (positive current through the unit from center pin to outer shell) for a few seconds. The duration of the current was not critical. After welding, the unit was sealed at the plug end with low melting solder or glyptal, if the crystal were to be used for other than test purposes.

In order to determine the force on the contact as a function of whisker deflection, or advance of the plug after contact, a standard pin with whisker welded in place was pressed into a hollow cylindrical aluminum rider. The assembly as shown in Fig. 2 was suspended over a polished crystal by a strip attached to a vertical microscope. As the microscope barrel was lowered by either coarse adjustment or vernier, the strip was lowered, and the weight of the pin and rider was transferred to the whisker, now touching the crystal surface. The deflections of the whisker under various weights, added to the rider in successive steps, were observed through the Filar eyepiece of the horizontal microscope shown. Weights were added with the whisker raised from the crystal, and care was taken to avoid exceeding the elastic limit of the whisker. Figure 3 shows a plot of force at contact *vs.* whisker deflection for several standard 0.0015″ whiskers. Breaks in the curve are probably caused by sliding at the contact.

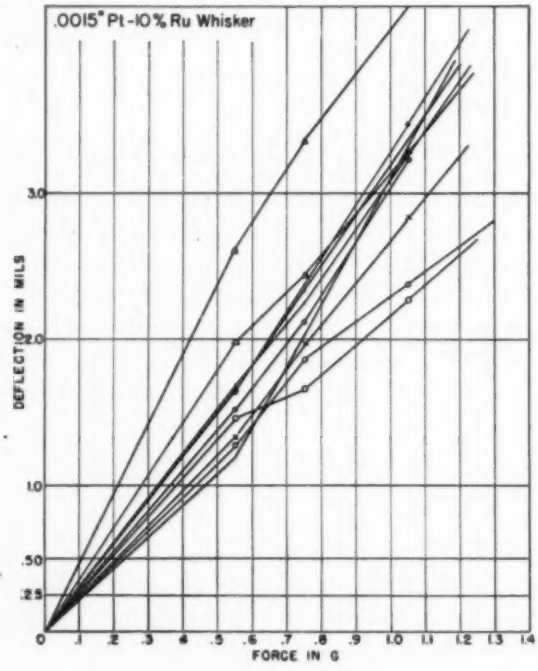


FIG. 3. Curves showing forces required to produce given deflections of standard 0.0015″ cat whiskers.

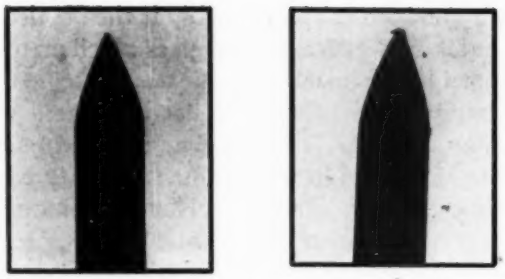


FIG. 4. (a) Sharpened whisker. (b) Whisker which was pulled away from a crystal after welding.

A series of microwave (3-cm wave-length) tests in cartridges had indicated that the best mixer performance occurred when 0.0005″ deflections were followed by a welding current of 250 milliamperes in the forward direction for a few seconds. Figure 3 shows that the 0.0005″ initial deflection corresponded to a force of about 150 milligrams. After passage of the welding current, an inverse deflection of about 0.0015″ or a force of about 500 milligrams was required to break the whisker loose. Figure 4(a) shows a sharpened whisker and 4(b), one which was pulled away from the crystal after welding. The diameter of the whisker below the germanium chip is approximately 0.0002″. Thus the welding current density through the contact was approximately 8.0×10^6 amperes per square inch. The tensile stress sustained by the weld was approximately 35,000 lb./sq. in.

The strength of the weld is again illustrated by the stability of a sealed cartridge with welded contact. Such a unit will stand at least a six-foot drop to a linoleum surface without suffering impairment of performance. Many silicon units of equivalent sensitivity will fail under such a test even though filled with the usual paratac-opal wax mixture to increase stability and to protect against humidity.

Efforts made elsewhere to weld silicon crystals have, to the author's knowledge, merely resulted in "burn out." This is probably caused by the greater chemical activity of silicon and to its oxidation in air. It is conceivable that silicon

crystals might be welded in an inert atmosphere if chemical combination with the whisker material is avoided.

4. THE D.C. CHARACTERISTIC

Let us consider briefly what d.c. current-voltage characteristic might be expected from a point rectifier. H. A. Bethe has shown that the natural barrier at the contact is thick enough to make diode theory applicable.[4] His treatment with simplifications can be connected satisfactorily to the phenomena observed in our welded units. Assuming a Maxwell-Boltzmann distribution of velocities for electrons crossing the barrier, he obtained an expression for the electron current density from semi-conductor to metal:

$$j=\tfrac{1}{2}\epsilon n v_0 \exp[-U/kT](\exp[\epsilon\varphi_a/kT]-1). \quad (1)$$

Here n is the density of electrons or donor atoms in the semi-conductor and v_0 is the average electron velocity.

$$v_0=(2kT/\pi m)^{\frac{1}{2}}; \quad (2)$$

ϵ is the electronic charge, k is the Boltzmann constant, and T is the absolute temperature. The factor $\frac{1}{2}$ occurs since only electrons moving toward the barrier get across it. In (1) φ_a is the potential in the forward direction applied to a barrier of height U above the potential in the metal. Bethe has considered the effect of image force upon the d.c. characteristic and has found it important in the case of unwelded contacts. If image force is neglected, $U=\epsilon\varphi_0$ where φ_0 is the contact potential or difference in work functions between the metal and semi-conductor. In this case (1) reduces to

$$j=\tfrac{1}{2}\epsilon n v_0 \exp[-\epsilon\varphi_0/kT](\exp[\epsilon\varphi_a/kT]-1) \quad (3)$$

or

$$I=I_0(\exp[\epsilon\varphi_a/kT]-1), \quad (4)$$

where I is the total forward current and I_0 is a constant. In addition to neglecting image force, this equation neglects the probability that electrons tunnel through the barrier. It has also been assumed implicitly that φ_0 is constant over the area of contact.

[4] This treatment has been described in reports having only limited distribution. A more detailed discussion will be found in a forthcoming book by H. C. Torrey and C. A. Whitmer of the M.I.T. Radiation Laboratory, *Crystal Rectifiers* (McGraw-Hill Book Company, Inc., New York, 1946).

In Eq. (4) the quantity φ_a is not one which can be measured directly. It is the potential applied to the barrier itself, and it does not include the voltage drop due to current passing through the series or spreading resistance of the crystal. The accepted equivalent circuit for a contact rectifier is shown in Fig. 5. Here R_s is the series or spreading resistance, R_x is the non-linear resistance because of the electronic barrier, and C_x is the capacity across this barrier. The spreading resistance is generally obtained by integration over the current path, assuming that the contact consists of a plane circular area between the flattened point and the semi-conductor. It is known that the resistance of such a contact is given by the expression,

$$R_s=\rho/2d, \quad (5)$$

where ρ is the resistivity of the semi-conductor and d is the diameter of contact.

Since φ_a in Eq. (4) is the potential applied to the barrier of the rectifier, one may re-write the equation as,

$$I_F=I_0(\exp[\alpha(V_F-I_F R_s)]-1). \quad (6)$$

Here V_F is the forward voltage applied to the rectifier and α has been written instead of the constant ϵ/kT. It is clear from Eq. (6) that if V_F is large compared to the term $I_F R_s$ in the exponent, and if the exponential is large compared to 1, a plot of log I_F *vs.* V_F should be a straight line. Furthermore its slope should be ϵ/kT or about 40 volt^{-1}. In general others have found that α is smaller than this and that it is not constant over a large range of currents. R. G. Sachs, H. J. Yearian, and others at Purdue University have explained these facts by assuming a variation of contact potential over the area of contact.[4] The welded contact microwave crystals follow an exponential law over a wide range of currents. The value of α obtained is usually more than 60 percent of the theoretical value.

Figure 6 shows a plot of the logarithm of the current against forward applied voltages for a

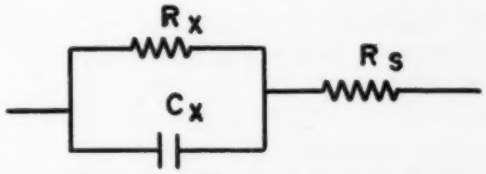


FIG. 5. Equivalent circuit for a contact rectifier.

typical welded contact germanium crystal and for a 1N31 silicon crystal taken at random from stock. The logarithm of the back current is also plotted against the applied back voltage for the two units. The forward characteristic for the germanium crystal is seen to be a straight line over more than three decades of current. The dotted curve is the continuation of the straight portion of the logarithmic characteristic. It actually represents a plot of the logarithm of the current as a function of the voltage applied to the barrier alone as the data points show. To obtain these points, values of R_s were tabulated which would create a voltage drop sufficient at each current to restore the solid curve to the dotted straight line. The average value of R_s was then multiplied by the current at each experimental point, and the voltage drop in the spreading resistance was subtracted from the solid experimental curve. The resulting points fell along the experimental curve as shown, illustrating the fact that the welded contact germanium crystal follows the theoretical expression over five decades of current. For the silicon unit no value of R_s could be found which would restore the experimental curve to a straight line as predicted by the simple theory. We have attributed this fact to the less intimate contact in the silicon unit. The value of α for the germanium unit was 24 volt^{-1}. For this unit I_0 was 9.05×10^{-5} milliampere. R_s had an average value of 2.59 ohms. Using Eq. (5), the measured resistivity of 0.005 ohm-cm for the germanium, and assuming a contact diameter of 0.0002", one obtains a value of 4.9 ohms. The discrepancy may be caused by the fact that the contact area in a welded unit is probably more nearly a hemisphere than a circular disk. If one makes the reasonable assumption that the current flow near the hemisphere is radial, one obtains the formula

$$R_s' = \rho/\pi d, \tag{7}$$

where R_s' is the spreading resistance of the welded unit, ρ is the resistivity of the germanium, and d is the diameter of the hemisphere. The same value of ρ and d used in Eq. (5) now yield $R_s' = 3.1$ ohms which is in better agreement with the measured value of 2.59 ohms. This low value of spreading resistance accounts in part for the straight portion of the log I_F *vs.* V_F plot.

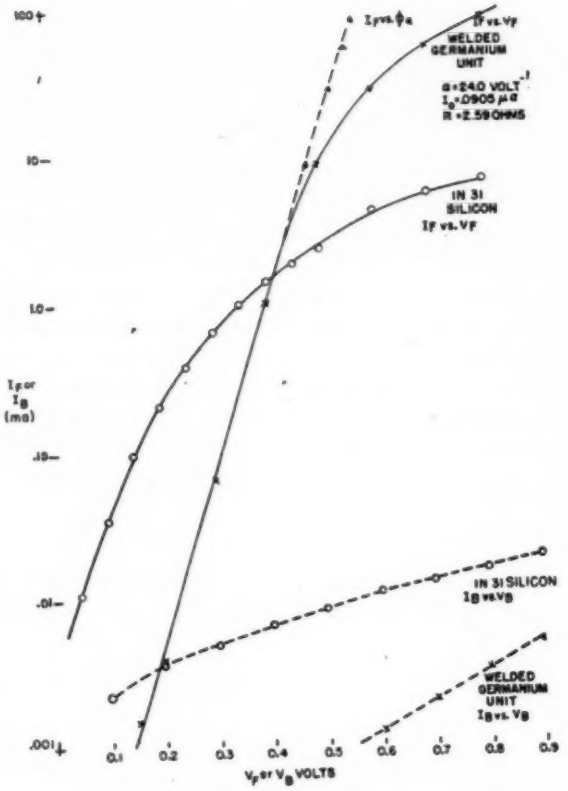


FIG. 6. Comparison of logarithmic characteristics.

Aside from its interest in connection with rectifier theory, the welded contact offers interesting possibilities in applications where resistors are required which follow a single exponential law over a wide range of current. In general the welded units offer such a characteristic over three decades of current.

5. DEFINITIONS OF CONVENTIONAL TERMS FOR MICROWAVE MIXER CRYSTALS

In discussing the microwave performance of mixer crystals three quantities are generally considered: The conversion loss and the noise ratio in turn determine the noise figure which combines the first two quantities. The following definitions are consistent with those proposed by H. T. Friis:[5]

1. Conversion loss, L_{db}.—Expressed in decibels, conversion loss of a crystal is defined by the equation,

$$L_{db} = 10 \log_{10} (P_s/P_\gamma), \tag{8}$$

where P_s is the maximum *available* power from the signal

[5] H. T. Friis, "Noise figures of radio receivers," Proc. I.R.E., **32**, 419–422 (1944).

source at frequency f_s, and P_γ is the maximum *available* power from the crystal at intermediate frequency γ. P_s is equal to the power *absorbed* by the crystal and P_γ is that *absorbed* by the i-f amplifier only when the crystal is matched to the wave guide or coaxial line, and to the i-f load. In general, if the crystal is mismatched to the incoming signal, L_{db} is greater than it is in the matched case since P_γ is reduced.

2. Noise ratio, N.—The noise ratio of a crystal is defined by the equation,

$$N = P_x/P_r, \tag{9}$$

where P_x is the maximum noise power *available* from the crystal to a noiseless resistor, and P_r is the maximum fluctuation noise power *available* from a resistor at room temperature to a noiseless resistor. Thus if the crystal creates no more noise than a resistor at room temperature, its noise ratio will obviously be unity. The quantity P_r is proportional to the product of the absolute temperature T of the resistor and to the band width, Δf, under consideration. It is given by the equation,[5]

$$P_r = kT\Delta f, \tag{10}$$

where k is the Boltzmann constant.

The noise ratio of a crystal can of course be expressed in decibels as well as the loss. So expressed, it is given by the formula,

$$N_{db} = 10 \log_{10} (P_x/P_r). \tag{11}$$

We shall have occasion to use it in this form.

Expressed in decibels, the noise figure of a crystal is defined by the equation,

$$F_{db} = L_{db} + N_{db}. \tag{12}$$

6. METHODS OF MEASURING LOSS AND NOISE RATIO

The early tests on microwave mixer crystals were made using a set simulating an actual radar receiver. This method of test is known as the "double detection" method. Local oscillator power at frequency $(f_s - \gamma)$ was fixed at a level to give 0.5 ma of rectified current through 100 ohms, and the noise power was observed at the output of the receiver. A known signal of frequency f_s was then introduced and increased until the power output of the receiver had doubled. Over-all receiver noise figure was obtained by dividing the known signal power required by $kT\Delta f$. Unfortunately the measurement is a very difficult one when applied to crystals because of the low power levels involved. Furthermore, it does not determine directly the conversion loss and noise ratio of the crystal alone.

To obtain the conversion loss of a crystal directly, a "modulation" method has been used and applied to the production testing of mixer crystals. R. N. Smith of Purdue University and S. Roberts of the Massachusetts Institute of Technology Radiation Laboratory have shown that modulation on a microwave local oscillator can be used to replace a low level signal in loss measurements. They have shown further that the voltage modulation on the r-f wave can be as high as 10 percent and that a 60-cycle modulation frequency is satisfactory. Loss values calculated from "double detection" results and those obtained by the "modulation" method showed excellent agreement. Roberts has shown that in the "modulation" method, conversion loss is given by the formula

$$L_{db} = 10 \log_{10} \frac{m^2 P_0}{V_L^2/R_L}. \tag{13}$$

Here m is the fractional r-f voltage modulation (60-cycle), P_0 is the average available oscillator power, and V_L is the 60-cycle (r.m.s.) voltage appearing across the load resistance, R_L.

Figure 7 shows a photograph of the microwave "plumbing" used for the loss measurement of the welded units. They were matched to the local oscillator. At the left is a 723-AB oscillator set to a wave-length of 3.2 cm. Following it is a variable resistor strip attenuator and a rotating resistor disk modulator. Fortunately an almost perfect sine-wave modulation is created by the modulator if an eccentric circular disk is used. This greatly simplifies the calculation of the fraction of modulation, m, which can be obtained by measuring maximum and minimum power passed by the modulator. To the right of the modulator in Fig. 7 is a resistor strip attenuator to decouple the modulator from the mixer and a standing wave detector to aid in matching the crystal to the wave guide. Figure 8 shows the d.c. and 60-cycle load circuit used to permit zero d.c. bias voltage regardless of the 60-cycle load. The battery potentiometer was varied until no direct current passed through the load. The load resistor was tapped to permit a variation of 60-cycle resistance as seen by the crystal. The a.c. voltage, V_L, across the load was measured with a Ballantine vacuum tube voltmeter.

The present measurement of noise ratio, although more difficult than that of loss, is simpler than the measurement of noise figure by the "double detection" method. The welded-contact

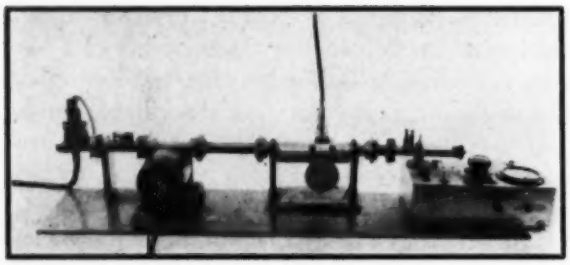


FIG. 7. Photograph of microwave plumbing used for loss measurement.

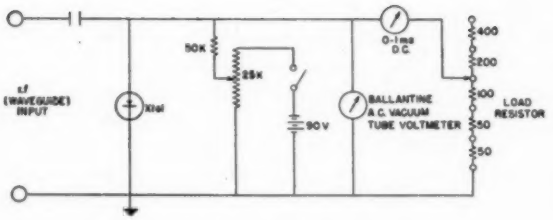


FIG. 8. Load circuit for loss equipment.

crystals described here were measured, in the matched case, on a standard noise set designed at the Massachusetts Institute of Technology Radiation Laboratory. The mixer was altered to permit matching each crystal to the padded wave guide. Since this noise set has been described in detail elsewhere,[6] its principles of operation will be merely outlined here. The set consists of a narrow-band 30-megacycle amplifier whose output meter reading is proportional to the power dissipated in its input resistance. An impedance matching circuit connects the amplifier input to the crystal under test. It is of such design as to permit the testing of crystals having i-f impedances between 150 ohms and 600 ohms. Using a known diode noise source, one may calibrate the output meter directly in terms of the noise ratio of the crystal at the input.

7. PERFORMANCE DATA ON WELDED MIXER CRYSTALS MATCHED TO THE LOCAL OSCILLATOR

Several hundred welded-contact germanium crystals were tested for loss by the modulation method and for noise by the method just described. In the loss measurement 1 milliwatt of local oscillator power was matched into the units. The crystals fed a 400-ohm 60-cycle load resistance. The d.c. bias was set to zero, although with 1 mw of available local oscillator power, bias was not critical. Figure 9 shows the loss distribution of a typical group of 177 welded crystals matched to the local oscillator. Figure 10 shows the noise distribution for the same group excluding 13 crystals which did not give sufficient current. During the noise measurements the units were matched to the local oscillator which was attenuated to give 0.5 ma of crystal current in each case. The values of loss and noise ratio for these units welded with 250 ma of d.c. are considerably lower than those obtained earlier for unwelded units. In general the crystals having low conversion loss are those with low noise ratios as well. Some improvement over the values of noise ratio plotted can be obtained by lowering local oscillator level. If bias is used, no appreciable increase in loss accompanies the decrease in noise.

[6] S. Roberts, 1N23 Noise Measuring Set Type 7438, Radiation Laboratory Report M-190 (Dec. 21, 1944).

8. CONVERSION GAIN IN WELDED UNITS MISMATCHED TO THE LOCAL OSCILLATOR

During the measurement of conversion loss of welded contact germanium crystals, it was noted that somewhat lower losses could be obtained when the crystals were mismatched to the local oscillator, creating a voltage standing wave ratio of about 5 to 1 in the wave guide ahead of the crystal. It was found that if an i-f load above 1000 ohms were used, the conversion loss in db often became negative or conversion gain occurred. Up to this time it had not been known that a crystal could furnish more power to the i-f load than it received as low level signal in the presence of a local oscillator. The Massachusetts Institute of Technology Radiation Laboratory, and the University of Pennsylvania, as well as the General Electric Company started groups working with the welded units furnished by the author, and the subsequent evaluation of the effect arose through the joint efforts of all concerned.

M. C. Waltz and R. H. Dicke of the Radiation Laboratory found that a resonant circuit attached to the i-f terminals of these units (fed by a mismatched microwave oscillator) would oscillate at the resonant frequency of the circuit, at least up to 34 megacycles. The presence of negative resistance was thus indicated. Figure 11(a) is a

photograph of the characteristic of a crystal with about 1 milliwatt of r-f applied but mismatched to the crystal. There are actually two regions of negative resistance observed, one near zero bias and one at a negative bias of about 0.9 volt.

To explain the conversion gain of the welded units, it is only necessary to explain the mechanism creating negative resistance at the i-f terminals. This can be seen most easily by a consideration of Fig. 11(b). It is a photograph of the characteristic of Fig. 11(a) when 60-cycle modulation (6 percent) was applied to the r-f. The blurred portion in the negative resistance region consists of a series of characteristics with r-f power as the parameter. Consequently a load line drawn through this region defines the limits of the i-f voltage swing for the modulation used. The load line actually drawn in Fig. 11(b) is nearly parallel to the negative resistance portion of the characteristic. Consequently a very small signal or r-f modulation will cause a large voltage swing along the i-f load line. This swing is almost independent of the r-f signal level. Thus, conversion gain can result. (The term gain is actually not uniquely defined when such a characteristic is obtained.) H. C. Torrey of the Radiation Laboratory has shown that with reasonable assumptions, linear network theory will account for the negative i-f conductance. His argument requires the variation of contact capacity with bias and low spreading resistance, both of which occur in welded germanium rectifiers.

W. C. Hahn of the General Electric Research Laboratory, using the assumption of a non-linear capacity set up a low frequency analogue to simulate the microwave performance of welded contact crystals. With it he was able to duplicate all of the phenomena observed with the microwave unit. The circuit producing the required characteristics is shown in simplified form in Fig. 12. This circuit approximates rather closely the one found in the microwave case. The resonant circuit including L and C exists in the wave guide mixer tuned to cause a large standing wave ahead of the crystal. L_w corresponds to the inductance equivalent to the combination of the glass bead and the cat whisker in the crystal cartridge. R_s corresponds to the spreading resistance of the germanium. R_x and C_x are the barrier

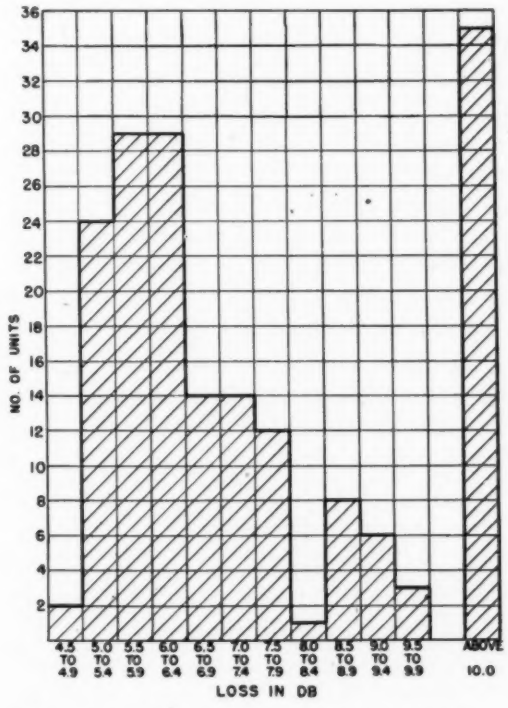


FIG. 9. X-band loss distribution for 177 welded crystals matched to local oscillator.

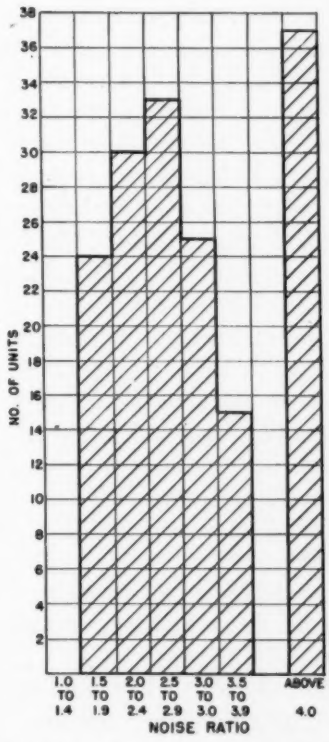
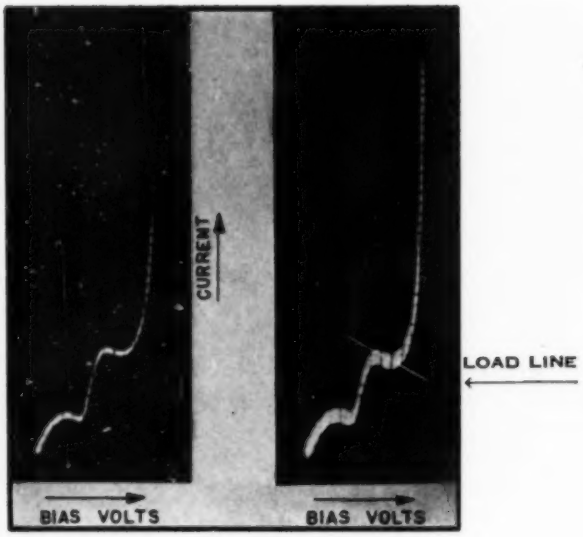


FIG. 10. X-band noise distribution for 164 welded crystals matched to local oscillator.



(a) (b)

FIG. 11. (a) Characteristic of a crystal with about 1 mw of r-f available but mismatched to crystal. (b) Photograph of the characteristic in (a) when 60-cycle moculation was applied to r-f.

resistance and capacitance varying with bias. C_x was obtained and varied in the low frequency analogue with a reactance tube circuit.

With the circuit of Fig. 12 Mr. Hahn readily obtained negative "i-f" resistances when R_s was sufficiently small and L_w was sufficiently large. The necessity for a low value for R_s agrees with the finding that silicon microwave crystals (having large spreading resistance) do not show negative i-f resistances. A few microwave crystals with short, L-shaped whiskers of decreased inductance were made. Tests confirmed Mr. Hahn's finding that L_w must be high, since no negative resistances were obtained. Rigorous analytical treatment of the circuit of Fig. 12 is very difficult. The general mechanism whereby conversion gain is obtained can be seen from the circuit, however. In the presence of modulation or r-f signal, the changing capacity, C_x, and resistance, R_x, throw the L-C circuit in the wave guide in and out of resonance. This varies the bias voltage at the modulation or intermedate frequency. The result is a transfer of local oscillator energy to the i-f terminals of the crystal.

It should be pointed out here that the negative resistances obtained are only possible because of the low spreading resistance and the circuit constants of the welded contact crystal. They are not caused by thermal effects as are the negative resistances in crystals heretofore described.[7] The conversion gain described for welded-contact germanium crystals depends upon the presence of a mismatched local oscillator, part of whose energy appears at the i-f terminals of the crystal when a signal is applied to the r-f terminals.

9. THE PERFORMANCE OF MISMATCHED CRYSTALS AS A FUNCTION OF BIAS

Since the welding process had made possible conversion gain in germanium crystals, a knowledge of crystal noise under conditions of mismatch became essential. For the measurement equipment on hand was used. The modulation set (Fig. 7) and the standard noise set were modified for the purpose. The equivalent circuit in Fig. 13 was assumed for the crystal as seen from its i-f terminals. Here i_n and i_s are constant current noise and signal generators for a given setting of the r-f tuning elements and a given d.c. bias. b represents the i-f susceptance of the crystal and g_{if} its i-f conductance. The circuit is commonly used in mixer representation and need not be justified here. The method of measurement was not conventional, however, and needs further discussion.

For the loss measurement mixer tuning was fixed on the modulation set to give the desired characteristic during the application of 1 milliwatt of available local oscillator power. The 1000-cycle admittance of the crystal to a low level signal was measured as a function of bias. It is known that if the r-f tuning elements are close to the crystal, its 1000 cycle admittance is equal, for all practical purposes, to its i-f conductance. The conductance, g_{if}, for a typical crystal has been plotted as a function of d.c. bias in Fig. 14. Note that the conductance becomes negative at a positive bias of 0.25 volt. To determine the value of i_s for a given signal input power, with the

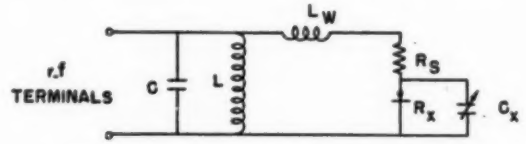


FIG. 12. Low frequency analog of a negative resistance crystal.

[7] E. C. Cornelius, "Germanium crystal diodes," Electronics **19**, 118–123 (1946).

mixer attached to the modulation set, the i-f load was set at 100 ohms. The 60-cycle voltage, V_L, across this load was measured as a function of d.c. bias. One milliwatt of local oscillator power was made available. From the value of V_L and g_{if}, i_s was determined for the equivalent circuit.[8] It may be seen from the circuit that the power *delivered* to the load conductance g_L is given by the expression

$$P_L=\frac{i_s^2}{(g_{if}+g_L)^2}g_L, \qquad (14)$$

when susceptance is negligible or tuned out. The usual definition of loss refers to power *available* from the i-f terminals of the crystal. When g_{if} is negative, power "available from the crystal" has no meaning. Consequently loss has been redefined for mismatched units as the ratio of available signal power to power *developed* in the load conductance, g_L. In Fig. 14, L_{db} is plotted as a function of d.c. bias voltage. For positive values of g_{if}, the usual definition of loss was used assuming $g_L=g_{if}$. A particular load of 0.0002 mho was considered in the negative resistance region since that is the order of grid-cathode conductances at intermediate frequencies.

The measurement of noise ratio was made on the standard noise set, but the input circuit was arranged as shown in Fig. 15. The equivalent crystal circuit of Fig. 13 has been included. The method used was essentially one of substitution. Crystal noise was compared with that of the temperature limited diode, D, shown in the figure. The inductance L_A, was used to tune out the capacitance of the diode, that of the 6AC7, and the susceptance of the crystal. The wave guide attenuator was set so that 1 milliwatt of local

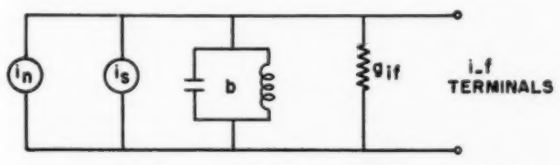


FIG. 13. Equivalent i-f circuit of a crystal rectifier.

[8] In measuring V_L in the region of low or negative i-f conductance, it is important either that the percent modulation be small enough or that the load line be steep enough to prevent distortion of the wave due to a large voltage swing at the i-f terminals. Six-percent modulation and a load of 100 ohms were found to be satisfactory.

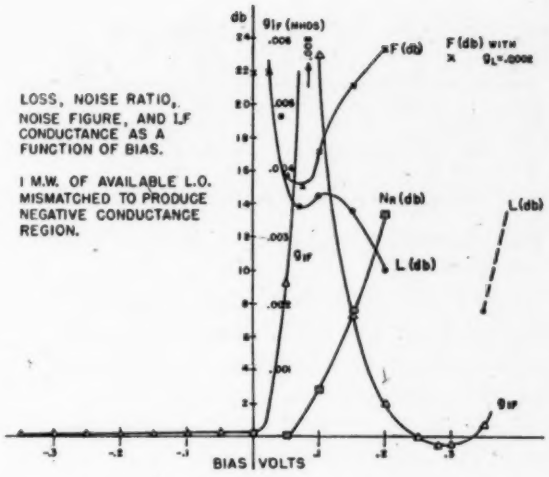


FIG. 14. Typical data on a welded crystal.

oscillator power was available as in the loss measurement. Wave guide tuning was left the same in the two measurements, so that i-f impedances were known for the noise measurement. The noise output for each crystal was compared with the known diode output, leaving the crystal in place when the diode was turned on, to insure the proper value of g_{if} in the measurement. In the negative resistance region, noise ratio was defined as the power *delivered* to the i-f load divided by $kt\Delta f$ (cf. Eq. (9)). For positive values of g_{if} the usual definition of noise ratio was used assuming $g_L=g_{if}$. In Fig. 14 N_{db} has been plotted as a function of bias. The 2000-ohm grid resistance of the amplifier prevented oscillation of the input circuit in the presence of the negative crystal resistance.

In Fig. 14, F_{db}, which is the sum of L_{db} and N_{db}, is also plotted. A continuous curve for this quantity has been plotted in the region of positive g_{if}. In the region of negative g_{if}, a typical value for F_{db} has been plotted from typical values of L_{db} and N_{db} in this region. For this crystal as for all others tested with mismatched local oscillator, N_{db} goes up as L_{db} goes down in such a way that F_{db} remains higher than its value obtained when the crystal is matched to the local oscillator.

The conditions under which these data were taken should be carefully specified. L_{db}, N_{db}, and F_{db} were obtained for the case in which signal, local oscillator and image were all mismatched alike. The modulation method was assumed applicable to these crystals. It should be remembered

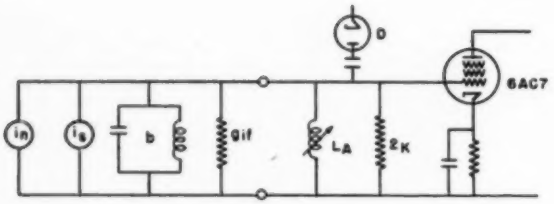


FIG. 15. Amplifier input circuit with crystal attached.

that as much as 3 db of signal power may be reflected from the mixer because of mismatch. The noise figures are in general so high, however, that this difference is not important. E. R. Beringer of the Radiation Laboratory has confirmed the results of the author by the "double detection" method, using a noise generator as a signal source. Although he tuned signal, local oscillator, and image separately, he too obtained no better performance than when the crystal was matched to the wave guide.

It thus appears from the measurements made on welded contact crystals, that the conversion gain obtained through mismatch at the r-f terminals is of little use in present radar receivers because of the increase in noise which accompanies the gain. Other schemes (e.g., regenerative circuits) for using the units have not yet been fully explored, however.

10. THE GENERATION OF HARMONICS

Because of the fact that microwave oscillators are not available for wave-lengths much below 1 cm, an investigation was made to determine the efficiency of welded contact germanium crystals compared, as harmonic generators, to 1N26 silicon crystals. Since standing wave detectors and bolometers were not available for the wave-lengths generated, absolute power measurements were not made. The work was undertaken since measurements made at the Massachusetts Institute of Technology Radiation Laboratory by C. G. Montgomery and D. D. Montgomery on our standard welded contact microwave crystals had indicated their superiority over silicon units as harmonic generators. To determine conditions for maximum harmonic output, contact force during adjustment and welding current were varied.

Figure 16 shows the wave guide arrangement used for comparing crystals as generators of the third harmonic of 1.2-cm radiation. The standard K-band mixer shown at the lower right of the figure was connected to a K-band padded wave guide from which 3 milliwatts of pulsed 1.2-cm power were available. The design of the holder is very similar to that used at the Radiation Laboratory. The d.c. terminals were accessible for biasing the generator crystal. Fundamental power was transmitted to the generator crystal shown at the upper right through the short section of coaxial line, choked to prevent escape of harmonic power back into the K-band guide. Four-mm power was carried by the 0.110" $\times$0.050" wave guide to the shielded detector crystal as shown in the figure. Tuning was done for the harmonic by adjusting plungers behind the generator and detector crystals and by sliding the generator crystal up and down on the spring finger contacts. Such tuning adjustments do not cover the entire range of possible impedances, but should suffice for the impedances usually encountered. During measurements the detector crystal was connected to the 750-ohm input of a 2000-cycle peaked amplifier. Thus the square wave (2000-cycle) modulated harmonic was detected and amplified for comparing crystals. Bias on the silicon detector crystal and r-f tuning were fixed for most sensitive operation. The generator crystal was matched to the fundamental source and to the harmonic guide in such a way as to obtain maximum output from the harmonic detector.

The harmonic output of a group of twenty

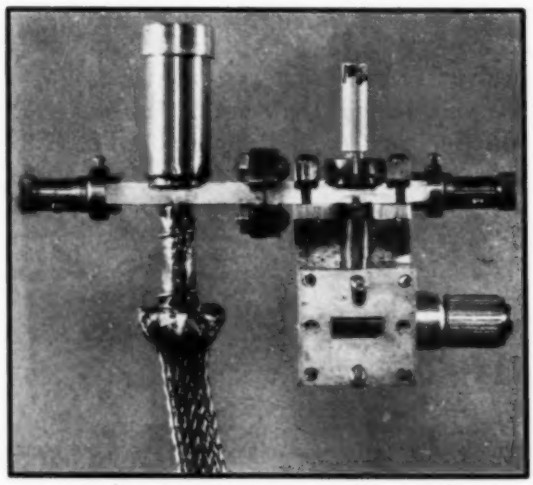


FIG. 16. Wave guide arrangement for comparing crystals as generators of third harmonic of 1.2-cm radiation.

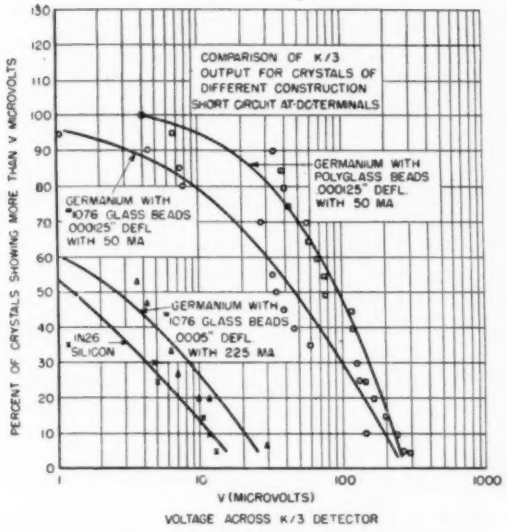

COMPARISON OF K/3 OUTPUT FOR CRYSTALS OF DIFFERENT CONSTRUCTION SHORT CIRCUIT AT DC TERMINALS
GERMANIUM WITH POLYGLASS BEADS .000125" DEFL. WITH 50 MA
GERMANIUM WITH "1076 GLASS BEADS .000125" DEFL. WITH 50 MA
GERMANIUM WITH "1076 GLASS BEADS .0005" DEFL. WITH 225 MA
1N26 SILICON
PERCENT OF CRYSTALS SHOWING MORE THAN V MICROVOLTS
V (MICROVOLTS)
VOLTAGE ACROSS K/3 DETECTOR

FIG. 17. Comparison of $K/3$ output for crystals of different construction.

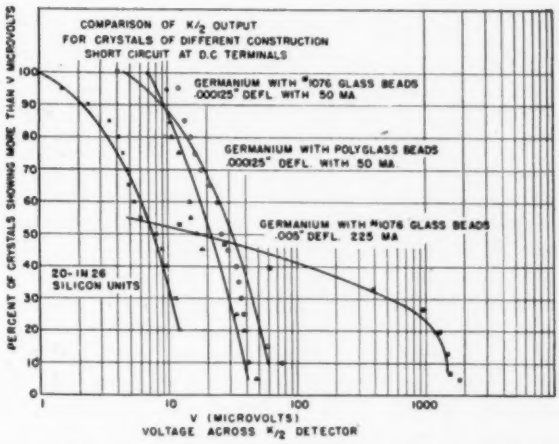

COMPARISON OF K/2 OUTPUT FOR CRYSTALS OF DIFFERENT CONSTRUCTION SHORT CIRCUIT AT D.C TERMINALS
GERMANIUM WITH "1076 GLASS BEADS .000125" DEFL WITH 50 MA
GERMANIUM WITH POLYGLASS BEADS .00025" DEFL. WITH 50 MA
GERMANIUM WITH "1076 GLASS BEADS .005" DEFL. 225 MA
20-1N26 SILICON UNITS
PERCENT OF CRYSTALS SHOWING MORE THAN V MICROVOLTS
V (MICROVOLTS)
VOLTAGE ACROSS K/2 DETECTOR

FIG. 18. Comparison of $K/2$ output for crystals of different construction.

1N26 silicon crystals was compared to that of each of three groups of welded germanium units. The first group was constructed and adjusted in approximately the same manner as the welded contact microwave units previously described. The 0.0005″ whisker deflection was followed by a welding current of 225 ma. The second group was the same except for the adjustment and welding current. The whisker deflection of this group was 0.000125″, and the welding current was 50 milliamperes. The third group of welded contact germanium crystals was like the second in adjustment and welding current, but the usual glass beads were replaced by beads of "Polyglass D," a thermoplastic material having a lower dielectric loss than the glass. The relative harmonic output of the three groups and that of the standard silicon units is shown in Fig. 17. The percentage of crystals showing more than V microvolts at the detector has been plotted against V, the detector voltage. It is apparent from the curves that our standard welded contact microwave crystals are somewhat better generators of 4-mm waves than the silicon units, and that more than twenty times the output of the silicon units can be obtained from welded germanium crystals with special contacts. Some improvement was noted when the "Polyglass" beads were used, presumably because of the decreased absorption of the harmonic.

Because of the possibility of obtaining much more second than third harmonic, the same groups of crystals were compared in a holder similar to that for the third harmonic. Figure 18 shows a comparison of the second harmonic output from the groups of crystals previously tested. It is interesting to note that the light-contact germanium units were only slightly better than the silicon units in second harmonic generation. It is also interesting that several of the germanium units welded in the usual manner showed very high second harmonic output.

11. ACKNOWLEDGMENTS

The author wishes here to thank Dr. A. W. Hull of the General Electric Research Laboratory for his encouragement throughout the project and for his suggestions concerning this presentation.

Thanks are extended to Dr. H. C. Torrey of the M.I.T. Radiation Laboratory, who first pointed out the linearity of the logarithmic characteristic of our units, and whose suggestions were extremely helpful during the course of this work.

To Miss M. E. Kershaw gratitude is expressed for her suggestions during the course of the work and for her patience in helping with many tedious operations and tests.

Fine Structure of Viscose Rayon*

H. G. INGERSOLL
Chemical Department, E. I. du Pont de Nemours & Company, Wilmington, Delaware
(Received May 31, 1946)

Quantitative x-ray diffraction methods have been used to study the orientation of the cellulose chains and lateral ordering around the long chain axes in various commercial and experimental rayons. Lateral order can vary independently of orientation in cellulose, as is the case with other linear polymers such as the polyamides. The degree of lateral order obtainable by aqueous annealing of regenerated cellulose appears to be restricted by the structure originally set up. As noted by Sisson, the well-known relations between yarn tenacity and orientation and elongation to break and orientation are not unique. At a given orientation, a range of tenacities and elongations are possible, depending upon the influence of other variables. Study of a series of yarns prepared under constant mechanical conditions showed that elongation to break increased with lateral disordering of the cellulose chains at constant orientation. The data support Baker's proposal that local chain disordering in any linear polymer favors higher extensibility. Some yarns of essentially the same orientation and lateral order have shown different properties indicating that other as yet undefined structural variables also have an important effect on physical properties.

INTRODUCTION

EXTENSIVE investigations over a period of years by Herzog, Hengstenberg, Meyer, Mark, Sisson, and many others have led to acceptance of the hypothesis that cellulose consists of both organized or crystalline and unorganized or amorphous regions.[1] The arrangement and dimensions, relative amount or perfection of the organized regions and the nature and arrangement of disordered regions vary widely depending upon the source and history of the sample. This heterogeneous order-disorder structure of cellulose, denoted by the term "fine structure," is believed to have an important effect on physical properties.

In this work, x-ray diffraction techniques have been applied for the quantitative measurement of parameters related to fine structure. Studies of relations among the various structural features and physical properties have also been undertaken.

EXPERIMENTAL TECHNIQUE

Equipment

The x-ray patterns were taken with a General Electric XRD unit using nickel-filtered copper radiation or in some cases monochromatized copper $K\alpha$ radiation. All patterns were taken with a 5-cm specimen-to-film distance. The films were processed under controlled conditions and the parameters were measured with a Photovolt photometer which had been remodeled and calibrated directly in x-ray intensity units.

Yarns Studied

With the exception of some saponified cellulose acetates, most of the samples studied were prepared by a conventional viscose process. Linters

TABLE I. Summary of x-ray parameters for viscose rayon.

Parameter	Description
	Orientation parameters—measured from azimuthal trace
$180/A$	A is the angular breadth in degrees of the interference at half maximum intensity, corrected for background intensity.
O, orientation number	Ratio of equatorial to polar intensity of the interference.
Orientation distribution	Angular distribution of the interference around the fiber axis, obtained by calculating the fraction of the total interference at each angle.
	Lateral order parameters—measured from radial trace
Reciprocal interference width (mm^{-1}) $(1/0.5W_2)$	Reciprocal of the half-width of the $10\bar{1}$ interference at half-maximum intensity, corrected for background intensity, measured toward the 101 interference.
Radial intensity Ratio $\frac{(I_1-I_m)}{I_1}100$	I_1—intensity of the 101 interference. I—minimum intensity between the 101 and $10\bar{1}$ interferences.

* Contribution No. 202 from the Chemical Department, Experimental Station.

[1] R. O. Herzog and W. Jancke, Zeits. f. Physik **3**, 196 (1920); J. Hengstenberg, Zeits. f. Krist. **69**, 271 (1928); H. Mark and K. H. Meyer, Zeits. f. physik. Chemie **B2**, 115 (1929); H. Mark, J. Phys. Chem. **44**, 764 (1940); W. A. Sisson, Chem. Rev. **26**, 196 (1940).

cellulose was steeped in sodium hydroxide, aged, then reacted with carbon bisulfide to form the soluble xanthate. The xanthate was dispersed in alkali then extruded into a primary bath which contained sulfuric acid and sodium sulfate and in some cases zinc sulfate or zinc and iron sulfates. The coagulated and partially regenerated filaments from the primary bath were then passed into a secondary bath containing hot water and stretched. Large differences in properties were achieved by variations in the viscose or baths.

X-Ray Parameters

Orientation

Figure 1 illustrates the x-ray diffraction pattern and photometric traces for a viscose yarn which was spun into an acid-sulfate bath and highly stretched. Parameters frequently measured from the patterns are tabulated and described in Table I and discussed in more detail in the following. The photometric trace may be made along the arc of an interference, giving the symmetrical intensity *vs.* angle curve shown,

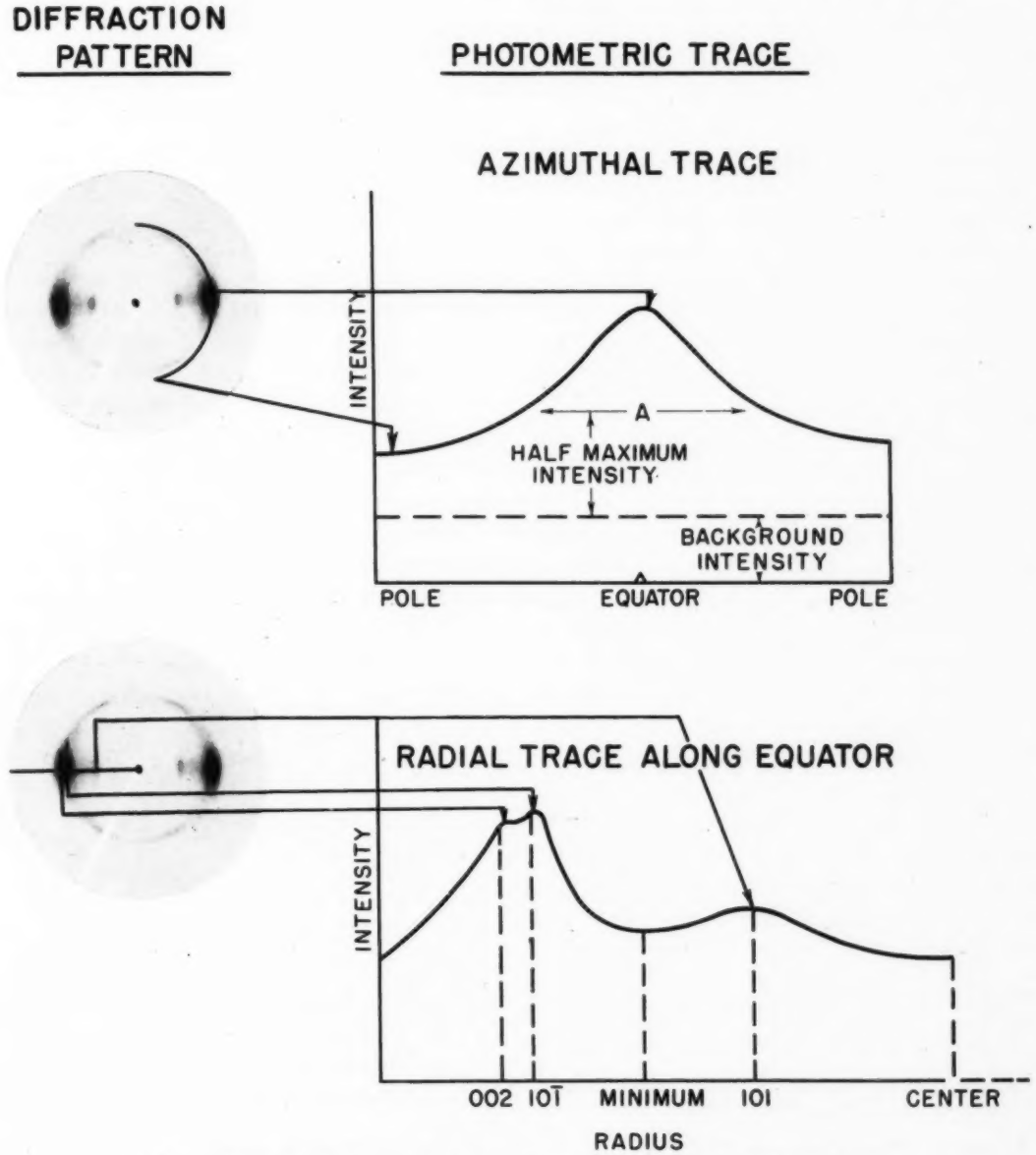


FIG. 1. Method of photometering x-ray diffraction patterns of viscose rayon.

which is a measure of the orientation of the cellulose crystallites. Complete orientation distributions can be calculated from this curve by procedures developed by Kratky, Hermans, Sisson, and others.[2] The distribution represents the fraction of the interference intensity having any designated orientation and is particularly useful in studying small differences in orientation or high degrees of orientation, as well as differences in orientation distribution. The angular breadth of the photometric curve at half-maximum intensity, a parameter used by Kratky, Sisson, and others,[3] is useful for measuring intermediate degrees of orientation and is more convenient to use than the complete distribution. We have expressed the measurements as 180° divided by the angle in order to provide a parameter which increases with orientation. We have also used another simple parameter, the ratio of the equatorial to the polar intensity, which is more suitable than the orientation angle for measuring low degrees of orientation.

Orientation of the long axes of cellulose crystallites parallel to the long axis of the fiber but with other crystallographic directions at random has been termed uniaxial orientation by Sisson.[4] In this case, equivalent orientation parameters are obtained from the three principal interferences for cellulose since all three interferences arise from planes parallel to the long axes of the molecules. Other types of orientation arise when other planes or directions in the cellulose crystallites are aligned with respect to particular reference planes or directions of the sample. A special type of orientation frequently encountered in cellulose is orientation of the 101 diffraction planes parallel to a surface of the sample, referred to as selective uniplanar orientation by Sisson.[5] In fibers this is manifested as orientation of the 101 planes parallel to the fiber surface (or axis). It is revealed by higher orientation parameters for the 101 interference than for the $10\bar{1}$ or 102 interferences. Unless otherwise specified the term orientation as used here refers to uniaxial orientation.

Background Correction

A troublesome experimental detail which should be mentioned is the background scattering correction which must be used in photometry because of the considerable diffuse scattering observed in patterns of regenerated cellulose. In many yarns, for example, the orientation angles and radial widths cannot be measured at all before a background correction has been applied. This has been discussed in detail by Kratky and others,[6] but no completely satisfactory solution has been arrived at. In the present work it was observed that cellulose yarns of low orientation frequently had a high degree of selective uniplanar orientation, that is, the 101 planes were more highly oriented than the $10\bar{1}$ or 002 planes. This selective orientation disappeared as the yarns were stretched and in highly oriented yarns the 101 planes actually appeared to be less highly oriented than the $10\bar{1}$ or 002 planes. There appears to be no physical reason for the 101 planes to develop a lower orientation than the other lateral planes at high uniaxial orientation. Consequently, it was assumed that after suitable background correction the orientation parameters measured from all three interferences should be equal in the highly oriented yarns. The minimum intensity between the 101 and $10\bar{1}$ interferences at 45° from the equator was chosen as an arbitrary correction factor since it was found to give about equal angles for all three interferences over quite a range of orientations. Although the structural significance of such a correction is uncertain, it serves the useful practical purpose of broadening the range of samples for which interference angles and half-widths can be measured.

Lateral Order

The photometric trace may also be taken radially along the equator of the pattern, Fig. 1, to obtain an intensity-distance curve representing the interferences arising from the three principal lateral spacings in crystalline cellulose. From this curve the intensities, positions, and

[2] O. Kratky, Zeits. f. Electrochemie **48**, 587 (1942); P. H. Hermans, Rec. trav. chim. des Pays-Bas **63**, 13–24 (1944); W. A. Sisson and G. L. Clark, Ind. Eng. Chem. Anal. Ed. **5**, 296 (1933).

[3] P. H. Hermans, O. Kratky, and R. Treer, Kolloid Zeits. **96**, 30 (1941); W. A. Sisson, Textile Research **7**, 425 (1937).

[4] W. A. Sisson, J. Phys. Chem. **40**, 343 (1936).

[5] W. A. Sisson, J. App. Phys. **44**, 523 (1940).

[6] P. H. Hermans, O. Kratky, and P. Platzek, Kolloid Zeits. **86**, 248–52 (1939).

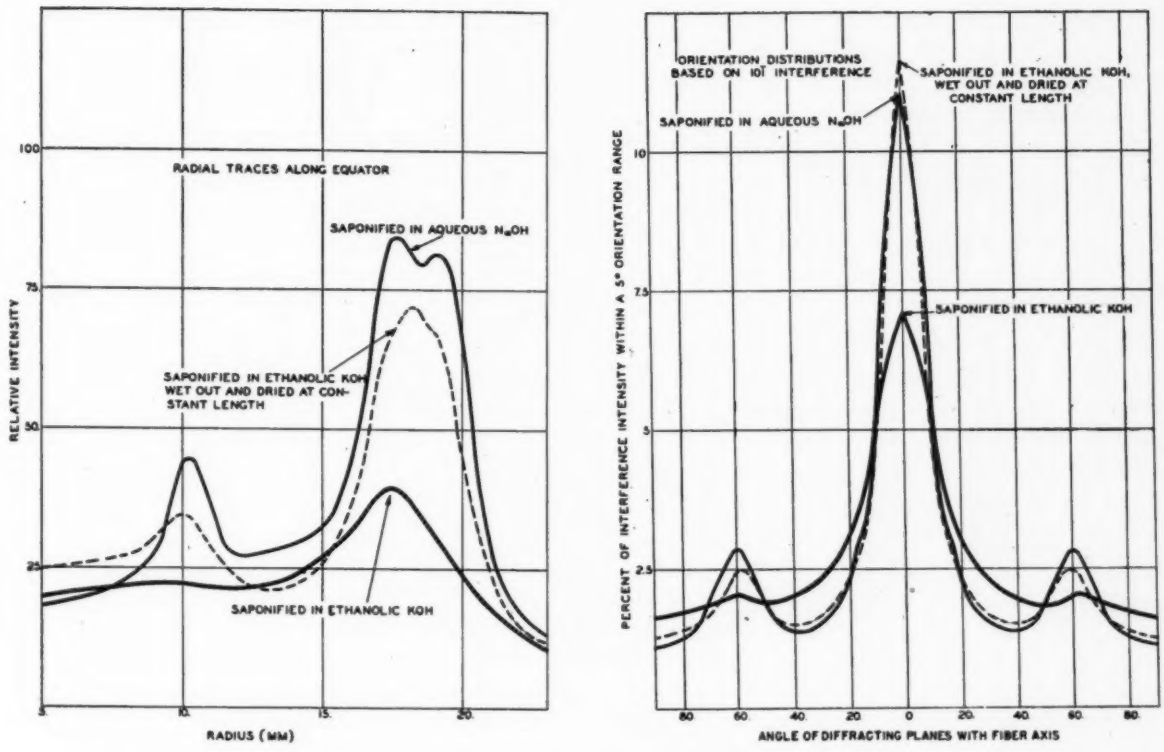

RADIAL TRACES ALONG EQUATOR
SAPONIFIED IN AQUEOUS N₂OH
SAPONIFIED IN ETHANOLIC KOH WET OUT AND DRIED AT CONSTANT LENGTH
SAPONIFIED IN ETHANOLIC KOH
RELATIVE INTENSITY
RADIUS (MM)
ORIENTATION DISTRIBUTIONS BASED ON 101 INTERFERENCE
SAPONIFIED IN ETHANOLIC KOH, WET OUT AND DRIED AT CONSTANT LENGTH
SAPONIFIED IN AQUEOUS N₂OH
SAPONIFIED IN ETHANOLIC KOH
PERCENT OF INTERFERENCE INTENSITY WITHIN A 5° ORIENTATION RANGE
ANGLE OF DIFFRACTING PLANES WITH FIBER AXIS

FIG. 2. Lateral order and orientation of cellulose yarns regenerated in aqueous and non-aqueous media.

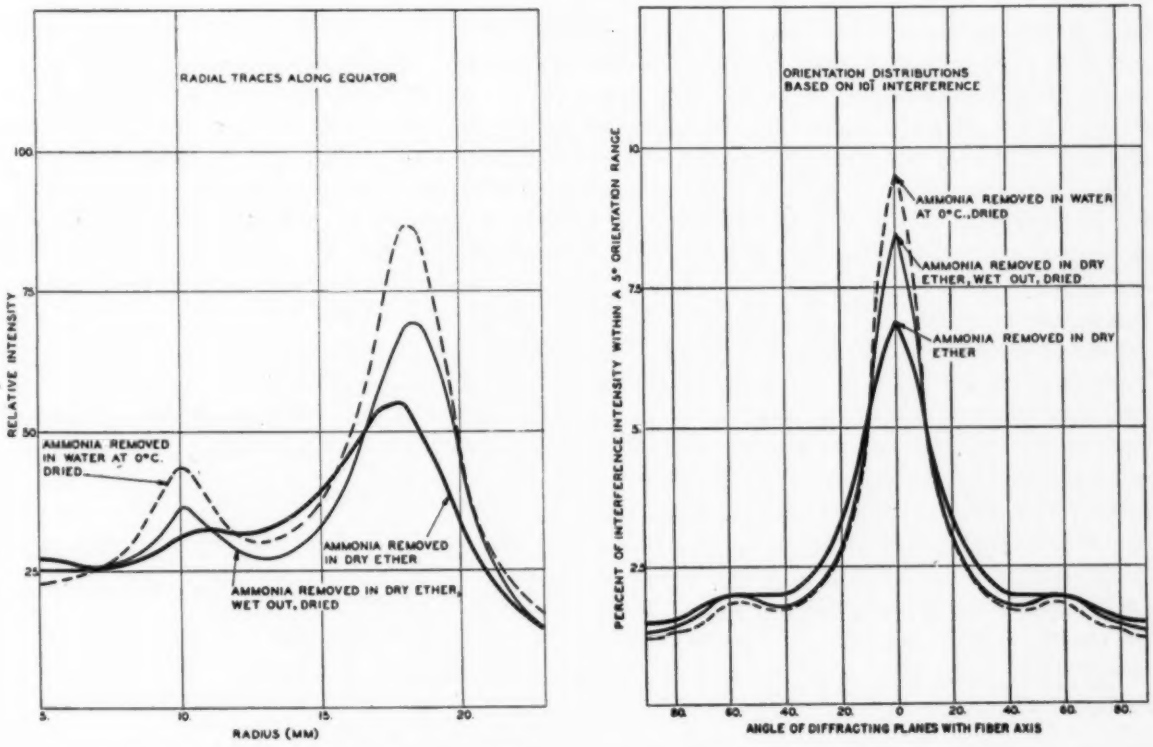

RADIAL TRACES ALONG EQUATOR
AMMONIA REMOVED IN WATER AT 0°C. DRIED
AMMONIA REMOVED IN DRY ETHER
AMMONIA REMOVED IN DRY ETHER, WET OUT, DRIED
RELATIVE INTENSITY
RADIUS (MM)
ORIENTATION DISTRIBUTIONS BASED ON 101 INTERFERENCE
AMMONIA REMOVED IN WATER AT 0°C., DRIED
AMMONIA REMOVED IN DRY ETHER, WET OUT, DRIED
AMMONIA REMOVED IN DRY ETHER
PERCENT OF INTERFERENCE INTENSITY WITHIN A 5° ORIENTATION RANGE
ANGLE OF DIFFRACTING PLANES WITH FIBER AXIS

FIG. 3. Lateral order and orientation of cellulose yarns after swelling-deswelling treatments.

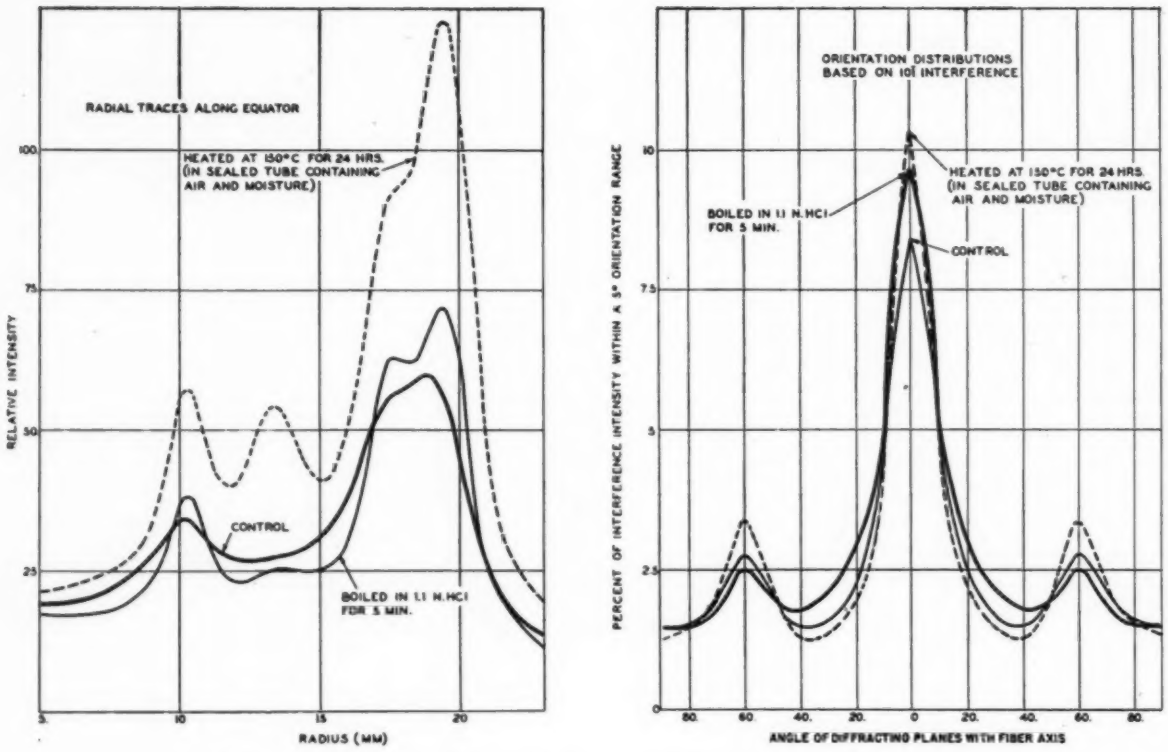

RADIAL TRACES ALONG EQUATOR
HEATED AT 150°C FOR 24 HRS. (IN SEALED TUBE CONTAINING AIR AND MOISTURE)
CONTROL
BOILED IN 1.1 N.HCl FOR 5 MIN.
RELATIVE INTENSITY
RADIUS (MM)
ORIENTATION DISTRIBUTIONS BASED ON 101 INTERFERENCE
HEATED AT 150°C FOR 24 HRS. (IN SEALED TUBE CONTAINING AIR AND MOISTURE)
BOILED IN 1.1 N.HCl FOR 5 MIN.
CONTROL
PERCENT OF INTERFERENCE INTENSITY WITHIN A 5° ORIENTATION RANGE
ANGLE OF DIFFRACTING PLANES WITH FIBER AXIS

FIG. 4. Lateral order and orientation of chemically degraded cellulosic yarns.

radial widths of the interferences can be measured. The widths, shapes, and intensities of the interferences are influenced by the dimensions and perfection of ordered regions and also by other fine structural features such as the amount, distribution, and arrangement of ordered and disordered regions. These structural features for the lateral planes depend largely upon the local disordering of the polymer chains in the direction perpendicular to their long axes, referred to as lateral disordering by Baker and Fuller.[7] This terminology is followed here, the width and intensity measurements being referred to merely as parameters related to lateral order in view of the uncertainty of more detailed structural interpretation for the very diffuse and overlapping interferences from some regenerated celluloses. The positions of the interferences are not used in this paper since the average interplanar spacings for the more ordered cellulose structures studied differ very little from those for quite disordered structures.

[7] W. O. Baker and C. S. Fuller, J. Am. Chem. Soc. **65** 1120 (1943).

The term lateral order is thus intended here to emphasize the fact that the x-rays give an overall picture of the degree of regularity of the lateral diffracting planes. Most of the regenerated celluloses here studied appear to be particularly unfavorable for attempts to assign specific structural causes such as variations in "degree of crystallinity" for the observed planar irregularity. Lateral ordering is of course to be differentiated from longitudinal ordering, i.e., ordering along the long chain axes.[7] Longitudinal ordering, which largely determines the sharpness of the non-equatorial interferences, is not dealt with in this paper.

EXPERIMENTAL RESULTS

Relation of Lateral Order to Orientation

Regenerating or Deswelling in Non-Aqueous Media

It has been noted that "crystallinity" as well as the orientation of cellulosic materials increases during stretching.[8] In considering the application

[8] P. H. Hermans, Proc. Acad. Wetenschappen Amsterdam **43**, 1032 (1940); Kolloid Zeits. **96**, 311 (1941).

of quantitative measurements of these features, the question then arises as to whether orientation may vary independently of lateral order. Sisson has observed that celluloses of very low order are produced by regenerating a cellulose derivative in non-aqueous media or by removing an intracrystalline swelling agent with a solvent less polar than water.[9] These effects have been examined quantitatively in order to indicate some of the possible relations between changes in lateral order and changes in orientation.

Figures 2 and 3 show photometric traces of diffraction patterns for cellulose yarns prepared from cellulose acetate by saponification in aqueous and non-aqueous media and for cellulose yarns after swelling in liquid ammonia and deswelling in aqueous and non-aqueous agents. The equatorial traces show that the interferences for the samples deacetylated or deswollen in aqueous media are very sharp and intense, indicating well-ordered diffracting planes. On the other hand, the interferences for the samples treated in non-aqueous media are weak and diffuse and the 101 interferences appear only as small inflections in the traces, showing that the cellulose chains are not sufficiently ordered to form well-defined diffracting planes. The chains are now nearly randomly turned about their long axes so that lateral separation of the chains is approximately the same in all directions, in contrast to the three quite different principal lattice spacings shown by a well-ordered cellulose. The orientations of the deswollen yarns, Fig. 3, are very little different from those of the original, which is the yarn of intermediate orientation in Fig. 7.

The samples prepared in non-aqueous media were unstable in that merely wetting and redrying markedly increased lateral order. Removal of the substituent groups or swelling agent in non-aqueous media leaves an irregular structure evidently because such media do not soften the cellulose sufficiently to allow ordering during the transition from the substituted or swollen network to the higher density cellulose. Wetting with water provides the necessary mobility and the chains rearrange, greatly increasing lateral order. The increase in order on wetting was accompanied by a marked increase in orientation in the case of the saponified cellulose acetate but there was not much change in orientation on wetting deswollen yarn.

It will be noted that wetting out the yarn deswollen in non-aqueous media converted it to a structure nearly as well ordered as that of the original untreated sample, shown as the yarn of intermediate orientation in Fig. 7. Comparable results were obtained by applying this swelling-deswelling and wetting out treatment to a yarn which was originally much more highly ordered as shown by the sharper radial trace in Fig. 7. The maximum degree of ordering achievable by drastic swelling followed by deswelling and aqueous annealing thus appears to be limited by the structure of the yarn originally set up during spinning. This point will be discussed further in a later section.

Mild Chemical Degradation

Another example of change in lateral order is furnished by chemical degradation. Mild chemical degradation of cellulose can greatly increase lateral order with substantially no change in orientation and no perceptible change in crystal

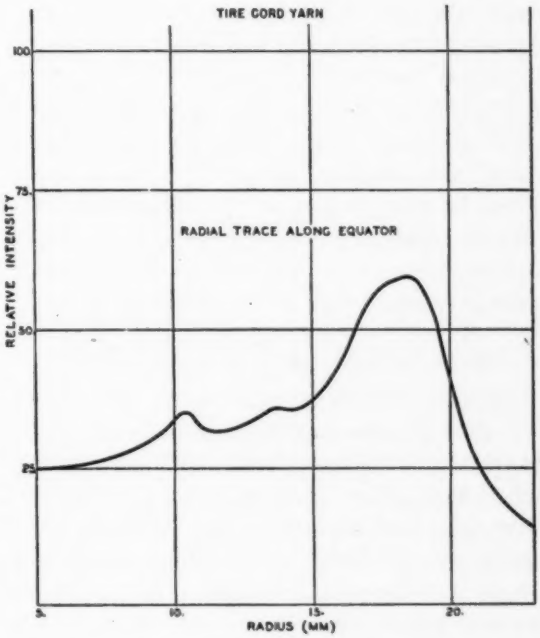


FIG. 5. X-ray diffraction from a yarn containing both mercerized and native cellulose crystal structures.

[9] W. A. Sisson, Ind. Eng. Chem. **30**, 530 (1938); W. A. Sisson, *Cellulose and Cellulose Derivatives* (Interscience Publishers, New York, 1943), first edition, Chapter IIIA.

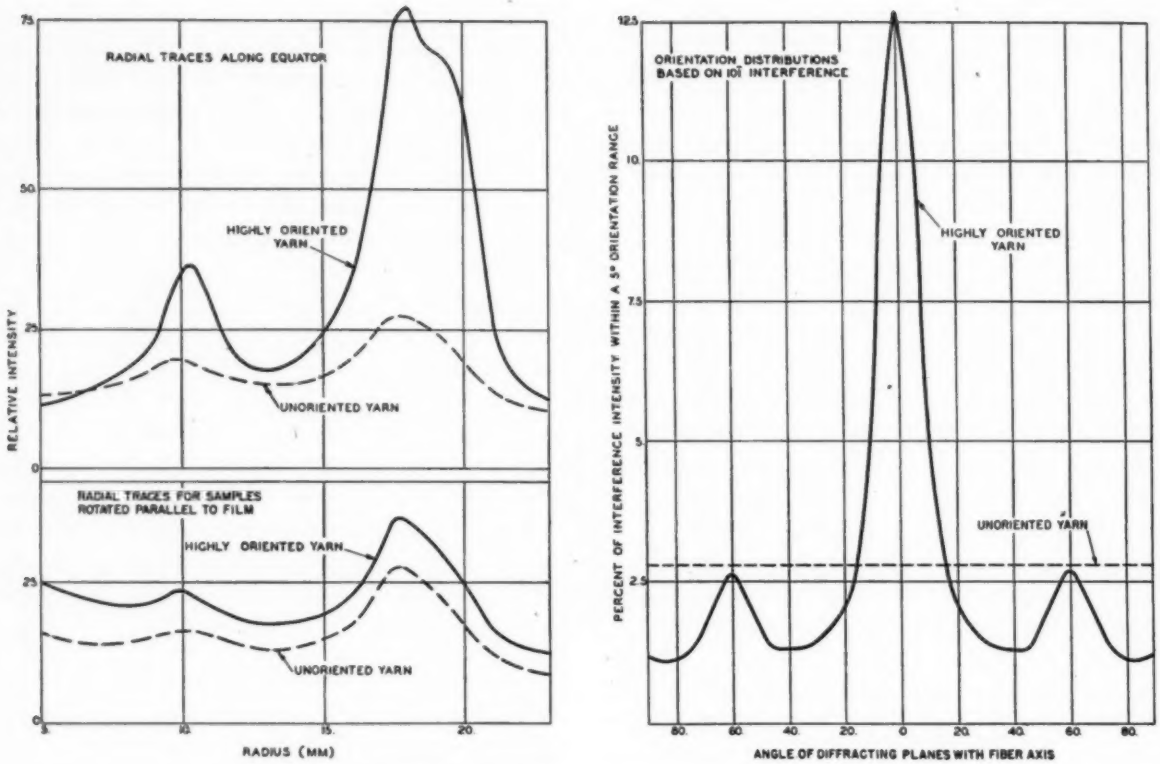

RADIAL TRACES ALONG EQUATOR
HIGHLY ORIENTED YARN
UNORIENTED YARN
RELATIVE INTENSITY
RADIAL TRACES FOR SAMPLES ROTATED PARALLEL TO FILM
HIGHLY ORIENTED YARN
UNORIENTED YARN
RADIUS (MM)
ORIENTATION DISTRIBUTIONS BASED ON 10$\bar{1}$ INTERFERENCE
HIGHLY ORIENTED YARN
UNORIENTED YARN
PERCENT OF INTERFERENCE INTENSITY WITHIN A 5° ORIENTATION RANGE
ANGLE OF DIFFRACTING PLANES WITH FIBER AXIS

FIG. 6. Effect of orientation on the measurement of lateral order.

structure. Figure 4 shows the photometric traces of a highly stretched tire cord yarn before and after 5 minutes boiling in 1.1 normal hydrochloric acid and after 24 hours heating in the presence of air and moisture. It is known that either of these treatments can cleave glucosidic linkages in the cellulose chains. This process occurring preferentially in the disordered, more accessible regions of the sample might release these less ordered chain segments sufficiently to allow rearrangement, and thus cause the over-all increase in lateral order which is observed.

Comparison of Cellulose with Other Polymers

The preceding examples show that lateral order can, to some extent, vary independently of orientation, a phenomenon which has been reported for other polymeric systems such as polyamides and cellulose esters by Baker.[10] The importance of distinguishing between changes in lateral order and orientation is emphasized. The increase in ordering on wetting the metastable cellulose obtained by regenerating or deswelling in non-aqueous media appears to be entirely analogous to the effects of annealing polyamides reported by Fuller, Baker, and Pape[11] and effects of heat and solvents on cellulose esters, described by the same authors.[12]

Lateral Order Parameters for Regenerated Cellulose

The differences in lateral order just illustrated can be expressed quantitatively in terms of the interference widths or intensities which can be measured from the photometric traces shown in Fig. 1. Unfortunately, the 101 interference is too diffuse in many rayons for accurate width measurement. The intensity of the 002 interference falls off rapidly, proceeding from the maximum outward along the radial trace, and furthermore, the position of the maximum is hard to determine because of the doublet formation with the 10$\bar{1}$ interference. Consequently the measured widths of the 002 interference vary little for large

[10] W. O. Baker, Ind. Eng. Chem. **37**, 246–54 (1945).

[11] W. O. Baker, C. S. Fuller, and N. R. Pape, J. Am. Chem. Soc. **62**, 3275–81 (1940).

[12] W. O. Baker, C. S. Fuller, and N. R. Pape, J. Am Chem. Soc. **64**, 776–82 (1942).

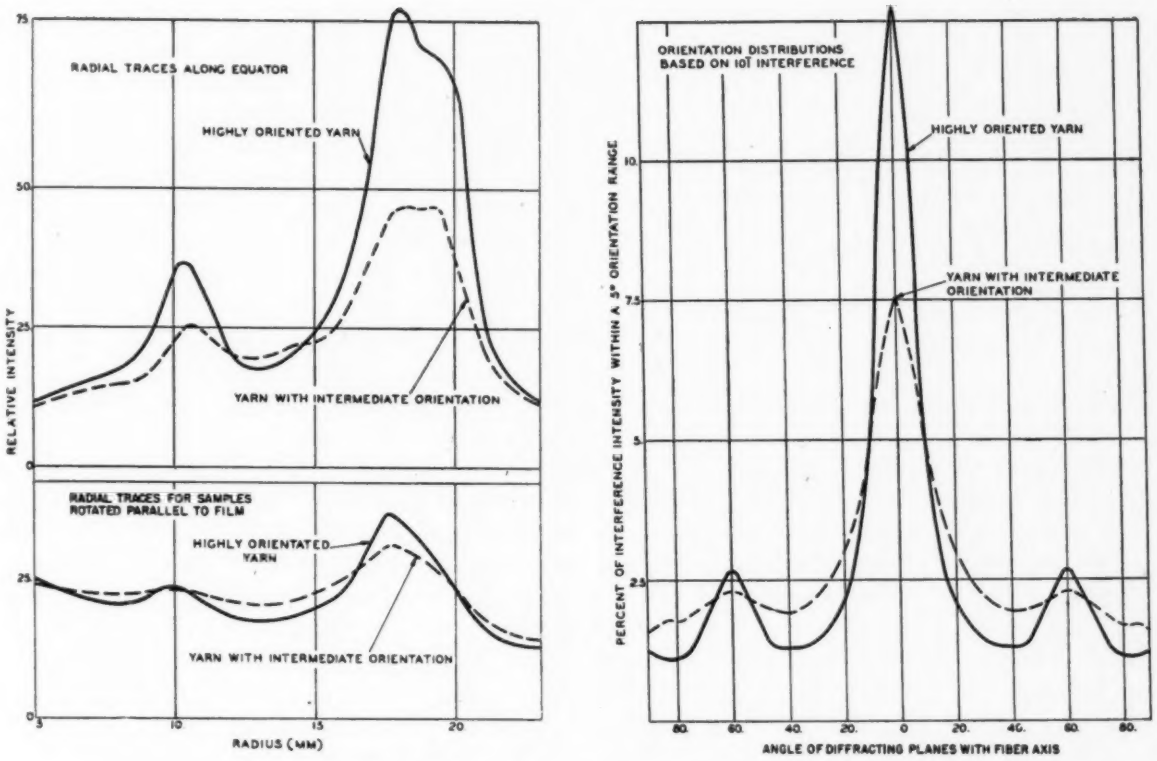

RADIAL TRACES ALONG EQUATOR
HIGHLY ORIENTED YARN
YARN WITH INTERMEDIATE ORIENTATION
RELATIVE INTENSITY
RADIAL TRACES FOR SAMPLES ROTATED PARALLEL TO FILM
HIGHLY ORIENTATED YARN
YARN WITH INTERMEDIATE ORIENTATION
RADIUS (MM)
ORIENTATION DISTRIBUTIONS BASED ON 10$\bar{1}$ INTERFERENCE
HIGHLY ORIENTED YARN
YARN WITH INTERMEDIATE ORIENTATION
PERCENT OF INTERFERENCE INTENSITY WITHIN A 5° ORIENTATION RANGE
ANGLE OF DIFFRACTING PLANES WITH FIBER AXIS

FIG. 7. Effect of orientation on the measurement of lateral order.

changes in pattern. The 10$\bar{1}$ interference width can be measured only toward the center of the pattern because of the 10$\bar{1}$−002 doublet formation and its usefulness is limited also by the fact that the native cellulose 101−10$\bar{1}$ doublet occurs just inside the 10$\bar{1}$ interference for regenerated cellulose, Fig. 5. Most rayon samples contain detectable amounts of native cellulose structure and many contain large proportions of it. A large difference in native cellulose content could change the 10$\bar{1}$ interference width even though there is no difference in lateral order. Nevertheless this parameter serves in many instances as a practical measure of pattern diffuseness. The width is expressed as a reciprocal in order to provide a parameter which increases with increasing lateral order.

X-ray intensities directly measured from the photographic patterns were not sufficiently reproducible at times because of variations in x-ray dosage. Consequently, an intensity ratio was adopted as an expedient for measuring differences in pattern sharpness more reproducibly. It was noted that the 101 interference almost disappears for celluloses of low lateral order and shows a large range of intensity depending upon sample history. The height of the 101 interference above the minimum between the 101 and 10$\bar{1}$ interferences, expressed as a percentage of the 101 intensity, $[(I_1-I_m)/I_1]100$, provides a parameter which increases with lateral order. This empirical ratio, which will be referred to as the radial intensity ratio, can be measured fairly conveniently and reproducibly by photometric methods.

Effect of Orientation on Lateral Order Parameters

It is emphasized that these lateral order parameters are only comparable if the samples have the same orientations. Small differences in orientation have little effect but large differences change the background corrections, interference intensities, and the contributions from non-equatorial interferences, with accompanying effects on the intensity ratios and reciprocal width parameters just described. In addition to this effect of uniaxial orientation, selective uniplanar orientation, that is, higher orientation of

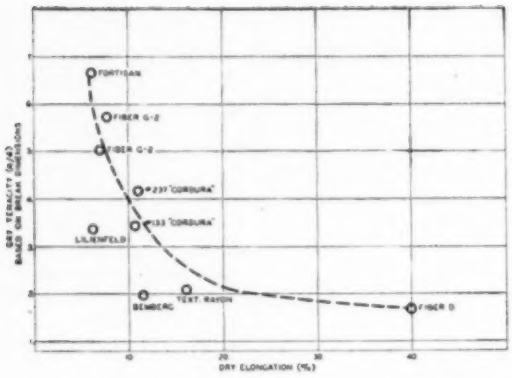


FIG. 8. Tenacity *vs.* elongation for standard rayons.

the 101 planes with respect to the $10\bar{1}$ or 002 planes, will increase the 101 intensity relative to the minimum. This produces an increase in intensity ratio even though there is no change in lateral order. Also, in samples containing appreciable amounts of material with the native cellulose crystal structure, special corrections must be made for the effect of native cellulose interferences on the intensity ratio. These latter two difficulties will be further discussed later as encountered in specific cases. To illustrate the effect of uniaxial orientation, Fig. 6 shows photometric traces of two yarns which have much different orientations and apparently different lateral order as revealed by the radial traces along the equator. If the yarn specimen is rotated in a plane perpendicular to the x-ray beam, during exposure, the interferences are distributed on uniform rings and the radial trace is equivalent to a radial trace averaged from equator to pole in the original unrotated patterns. The average traces, obtained from rotated samples, show that these two yarns actually have quite similar lateral order, the difference originally noted arising largely from the difference in orientation. Figure 7 reproduces the traces for samples differing widely in orientation and also apparently in lateral order, when the estimate is made from the original differently oriented samples. X-ray diagrams of these same yarns, averaged by rotating the samples, demonstrate that they are still quite different with respect to lateral order, even when compared at the same orientation.

It should be noted that rotating the samples does not completely average the diffraction, since the orientations of the various planes are not randomized by rotating around one axis. This discrepancy in comparing the lateral order is of course more important where the differences in orientation are largest. An attempt was made to randomize the samples completely before x-raying by cutting into lengths roughly equal to their diameter. However, a serious difficulty was encountered here in that it was difficult to be sure that the cutting had not altered fine structure. It was found, for example, that room temperature cutting considerably decreased lateral order as did also rolling dry cellulose filaments. This effect appears to be closely comparable to the disordering of cellulose derivatives by rolling described by Baker[10] and to the disordering of cellulose during grinding reported by Hess and co-workers,[13] indicating that mechanically working an unswollen cellulose decreases its lateral order. Cutting the cellulose at very low temperatures, such as dry ice temperature, appears to reduce the concurrent changes

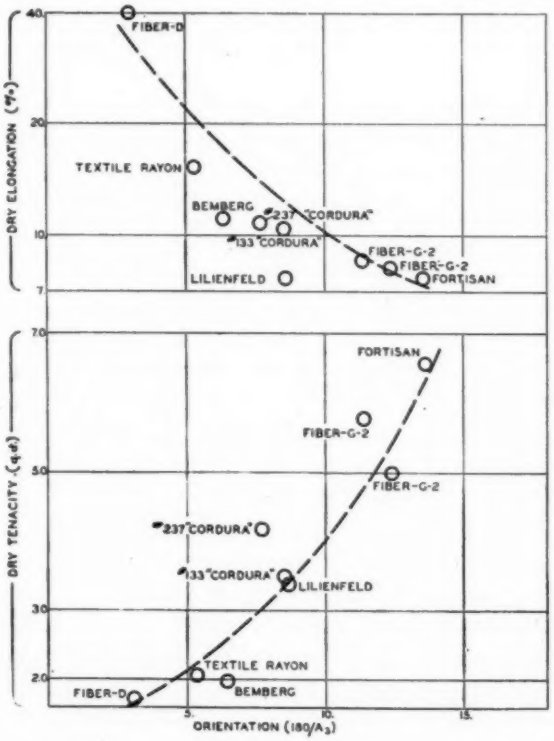


FIG. 9. Tenacity-elongation-orientation relations for standard rayons.

[13] K. Hess, E. Steurer, and H. Fromm, Kolloid Zeits. [1] **98**, 148–159 (1942); [2] **98**, 290–304 (1942).

in fine structure, but it is not certain that even this method of randomization is satisfactory.

These averaging procedures tend to increase the relative background scattering intensity and thus add to the difficulties of measuring small differences. However, the results show that a comparison of the lateral order of samples differing widely in orientation can be very misleading unless some correction is made for the difference in orientation.

Correlation of Properties with Fine Structure

The structural features, orientation, and lateral order, are considered to have important effects on physical properties of cellulose. It is well known from the work of Herzog, Lilienfeld, Berkley, Morey, Sisson, and others that many physical properties, including tenacity and elongation, vary in a regular way with orientation.[14] Hermans states that the properties and quality

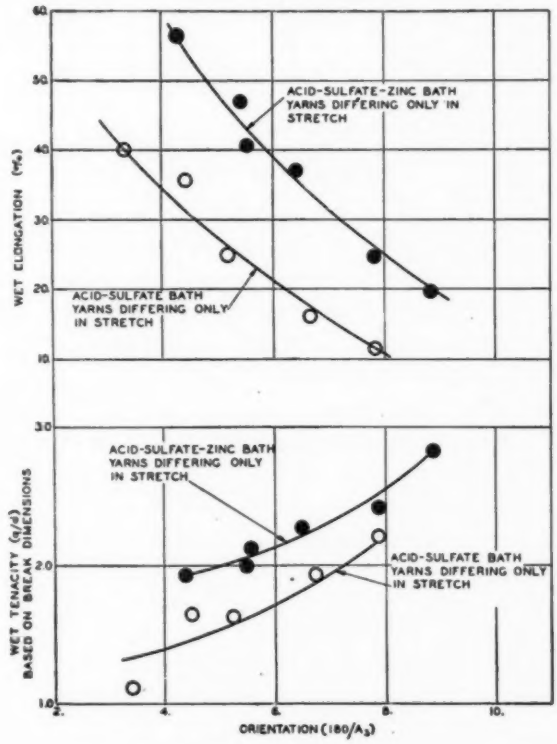

WET ELONGATION (%)
ACID-SULFATE-ZINC BATH YARNS DIFFERING ONLY IN STRETCH
ACID-SULFATE BATH YARNS DIFFERING ONLY IN STRETCH
WET TENACITY (g/d) BASED ON BREAK DIMENSIONS
ACID-SULFATE-ZINC BATH YARNS DIFFERING ONLY IN STRETCH
ACID-SULFATE BATH YARNS DIFFERING ONLY IN STRETCH
ORIENTATION (180/A₃)

FIG. 10. Wet tenacity-wet elongation-orientation relations for experimental rayons.

[14] R. O. Herzog, J. Phys. Chem. **30**, 457 (1926); L. Lilienfeld, Brit. Patents 274,521 (Jan. 11, 1926); 274,690 (Jan. 11, 1926); E. E. Berkley, Textile Research **9**, 355 (1939); D. R. Morey, Textile Research **4**, 491 (1934); W. A. Sisson, Textile Research **7**, 425 (1937).

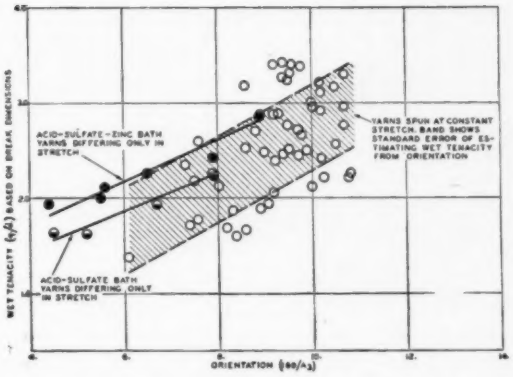

WET TENACITY (g/d) BASED ON BREAK DIMENSIONS
ACID-SULFATE-ZINC BATH YARNS DIFFERING ONLY IN STRETCH
YARNS SPUN AT CONSTANT STRETCH. BAND SHOWS STANDARD ERROR OF ESTIMATING WET TENACITY FROM ORIENTATION
ACID-SULFATE BATH YARNS DIFFERING ONLY IN STRETCH
ORIENTATION (180/A₃)

FIG. 11. Wet tenacity *vs.* orientation for experimental rayons.

of regenerated fibers depend upon (1) the degree of polymerization and distribution of chain length; (2) the preferred orientation of the crystallites and molecules and the distribution of this orientation; (3) the ratio of crystalline to non-crystalline material; and (4) the average distribution of the above over the cross section of the fiber.[15]

Mark has proposed that flexibility and reactivity are dependent upon the disordered regions in cellulose while tenacity and elastic modulus are related to the amount of ordered material present.[16] In the present work quantitative x-ray measurements have been applied to ascertain whether this effect of disorder on properties can be verified experimentally.

Tenacity—Elongation—Orientation Relations for Standard Rayons

As a broad introduction to the data, Fig. 8 shows tenacities plotted against elongations for a number of commercial types of cellulose yarns. Tenacity decreases with increasing elongation in such a way that a smooth curve averaging the points has a hyperbolic shape. This relation is well known, being indicated by the rough constancy of the product of tenacity and elongation which has been called the "silk factor" or "work factor."[17]

This tenacity-elongation relationship for yarns differing widely in orientation results from the fact that both tenacity and elongation vary with

[15] P. H. Hermans, J. Phys. Chem. **45**, 827 (1941).
[16] H. Mark, J. Phys. Chem. **44**, 764 (1940).
[17] D. Smith, Proc. A.S.T.M. **44**, 543 (1944).

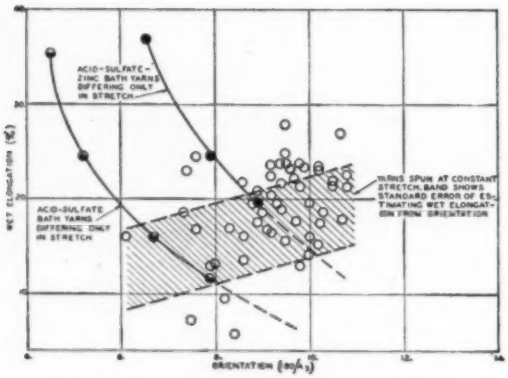

ACID-SULFATE-ZINC BATH YARNS DIFFERING ONLY IN STRETCH
ACID-SULFATE BATH YARNS DIFFERING ONLY IN STRETCH
YARNS SPUN AT CONSTANT STRETCH, BAND SHOWS STANDARD ERROR OF ESTIMATING WET ELONGATION FROM ORIENTATION
WET ELONGATION (%)
ORIENTATION (mμ²)

FIG. 12. Wet elongation *vs.* orientation for experimental rayons.

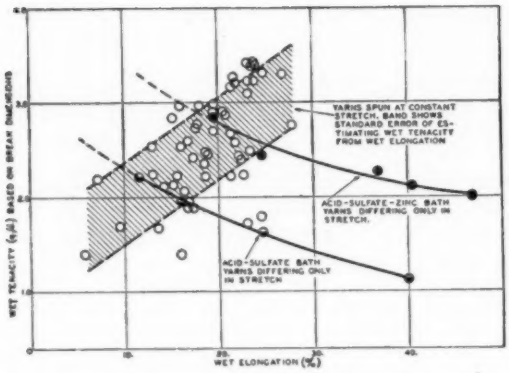

YARNS SPUN AT CONSTANT STRETCH, BAND SHOWS STANDARD ERROR OF ESTIMATING WET TENACITY FROM WET ELONGATION
ACID-SULFATE-ZINC BATH YARNS DIFFERING ONLY IN STRETCH.
ACID-SULFATE BATH YARNS DIFFERING ONLY IN STRETCH
WET TENACITY (g/d) BASED ON BREAK DIAMETERS
WET ELONGATION (%)

FIG. 13. Wet tenacity *vs.* wet elongation for experimental rayons.

the orientation of cellulose chains. This dependency is illustrated in Fig. 9 where both tenacity and log elongation are plotted against orientation for these yarns. Tenacity appears to increase substantially linearly with increasing orientation while elongation decreases hyperbolically with increasing orientation. The increase in tenacity with orientation has been explained qualitatively by Mark on the basis that the more highly oriented materials have more crystallites aligned parallel to the fiber axis and hence in a position to support the load.[18] Elongating the filaments increases the orientation so that if all samples break at complete or comparable orientations, more elongation would be possible in a less highly oriented sample. Preston has calculated elongation on this basis from refractive index measurements of orientation and obtained qualitative agreement with the observed values.[19]

The scatter of the data in Fig. 9 indicates that factors other than orientation also affect these properties. The curves shown have been drawn to represent the highest tenacities and elongations so far observed at each orientation. From the measured orientation of a given yarn and these reference curves, one can estimate the tenacity and elongation which the yarn should develop if it is as efficient as the reference yarns in converting orientation into these physical properties. The ratios of estimated to observed properties, which may be termed "orientation efficiencies," serve as a convenient indication of the effects of factors other than orientation and have been helpful in evaluating new yarns and spinning processes. In some instances, for example, low efficiencies have been traced to mechanical non-uniformities of the yarns. In other cases it has been possible to show that some structural features other than orientation were important.

Tenacity—Elongation—Orientation Relations for Experimental Rayons

Wet tenacity-orientation, wet elongation-orientation curves for two experimental yarns at various degrees of stretch are shown in Fig. 10. One yarn was spun into an acid-sulfate bath, the other into an acid-sulfate-zinc bath. Dry tenacities and elongations plotted against orientation for these same yarns at various stretches give curves which are quite similar but in general not as smooth as those shown for the wet properties. Differences between dry and wet properties, the treatment of which is beyond the scope of this paper, may be neglected for the purposes of this discussion. The linear tenacity-orientation and hyperbolic elongation-orientation relation indicated roughly by data for several types of rayon in Fig. 9 are shown more clearly here for these yarns in which the orientation is increased by merely increasing spinning stretch. The curves for these two experimental yarns have much the same form and slopes but are displaced from each other because of the effects of variables other than orientation. Changes in bath or viscose variables other than adding zinc

[18] H. Mark, Melliand Textiber **10**, 695 (1929); Pap. Trade J. **113**, 34 (July 17, 1941).
[19] J. M. Preston, *Modern Textile Microscopy* (Emmott and Company, London, 1933), first edition, p. 265.

to the bath have produced a similar effect, forming families of tenacity-orientation and elongation-orientation curves, respectively. At any given orientation, then, a range of tenacities and elongations is possible, depending upon other factors, as shown by Sisson.[9]

Some information on the effects of fine structural factors other than orientation was obtained by examining a series of yarns spun by identical mechanical procedures. Large changes in properties of these yarns were achieved by changing the acidity or metal ion concentration in the coagulating bath or the degree of substitution, salt index, or composition of the viscose. All samples were collected on a bobbin and dried at constant length. These changes in spinning conditions and viscose produced large differences in lateral order and marked differences in orientation, even though secondary bath stretch was held constant. The graph of wet tenacity against orientation for these yarns, Fig. 11, shows that the data scatter considerably, suggesting that variables other than orientation are important. There is, nevertheless, a trend for wet tenacity to increase with orientation and the over-all linear correlation coefficient calculated for this relation has high statistical significance. The shaded band shown represents the standard error of estimating wet tenacity from the best straight line expressing tenacity as a function of orientation, calculated by the method of linear regression.[20] The increase in wet tenacity with orientation brought about by stretch alone shown in Fig. 10, is also reproduced in the solid curves of Fig. 11 for comparison. Differences in orientation apparently have similar effects on tenacity whether achieved by variations in viscose and bath or by differences in degree of stretching.

With elongation, however, the situation appears superficially to be quite different from that encountered when stretch is varied. Elongation now appears to increase somewhat with orientation, Fig. 12. The solid curves show for comparison the wet elongation-orientation relations previously observed for yarns stretched different amounts, Fig. 10. The scatter of this wet elongation-orientation plot is also very large, again indicating that variables other than orientation are important.

Since both wet tenacity and wet elongation increase with orientation for these yarns spun at constant stretch, it follows that wet tenacity must also increase with wet elongation, as illustrated in Fig. 13. The solid curve represents the previously shown hyperbolic wet tenacity-wet elongation curve for yarns differing mainly in orientation. The over-all increase in wet tenacity with wet elongation encountered here suggests that the stress-strain curves may have generally similar shapes, the yarns merely breaking at different elongations. Obviously some specific yarns as shown on the slide must have considerably different stress-strain curves because of the observed large spread in elongation at a given tenacity or in tenacity at a given elongation.

Statistical Analysis of the Correlations among Structure and Property Parameters

In general, there appear to be relations among wet tenacity, wet elongation, lateral order, and orientation of these yarns, but the data scatter considerably, indicating that variations in other factors are important. Mathematical procedures have therefore been applied for a quantitative study of the complex interrelationships among the variables. Applying the usual statistical methods, linear correlation coefficients were calculated for each of the six possible pairs of variables, Table II. Statistical significance of

TABLE II. Total and partial linear correlation coefficients for properties and x-ray parameters of experimental rayons.

$n = 55$

Variables					
Wet tenacity (based on break dimensions)				*Tow*	
Wet elongation				*Ew*	
Orientation $(180/A_3)$				*O*	
Lateral order $[(I_1 - I_\infty)/I_1]100$				*L*	
Variables correlated	Total coefficients	"t" factors	Variables held constant for partial coefficients	Partial coefficients	"t" factors
Tow−*O*	+0.5312*	4.48	*Ew*, *L*	+0.4034*	3.15
Ew−*O*	+0.3910*	3.04	*Tow*, *L*	+0.0804	0.58
Tow−*Ew*	+0.5891*	5.21	*O*, *L*	+0.4310*	3.41
Ew−*L*	−0.5496*	4.71	*Tow*, *O*	−0.4492*	3.59
Tow−*L*	−0.3478*	2.65	*Ew*, *O*	−0.0133	0.09
L−*O*	−0.2595	1.92	*Tow*, *Ew*	−0.0477	0.34
"t" factor required for significance					
5% level				1.96	
1% level				2.58	

* Coefficient considered statistically significant.

[20] M. J. B. Ezekial, *Methods of Correlation Analysis* (John Wiley and Sons, New York, 1941), second edition.

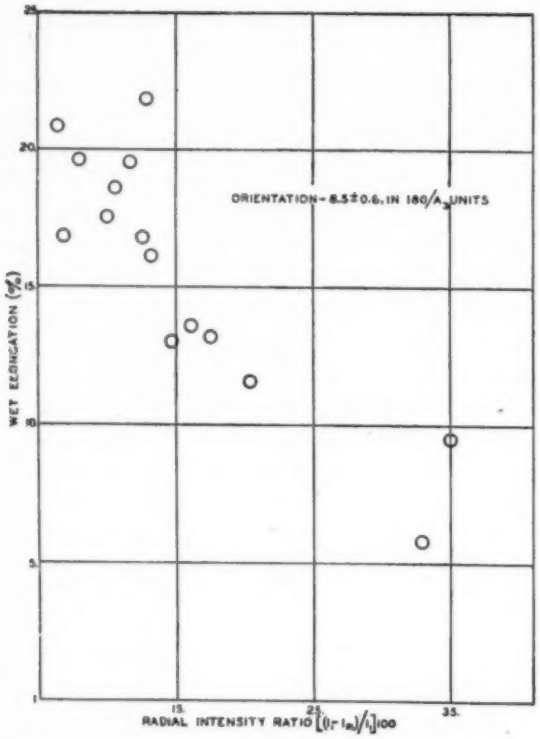


FIG. 14. Wet elongation *vs.* lateral order for experimental rayons of constant orientation.

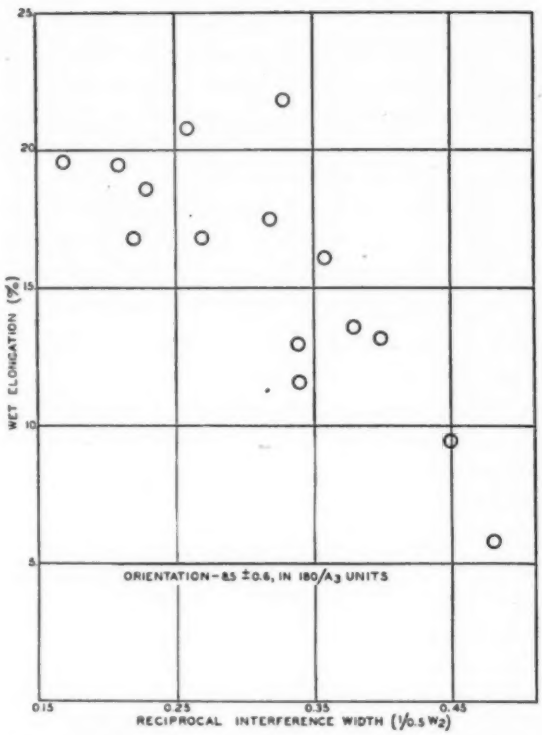


FIG. 15. Wet elongation *vs.* lateral order for experimental rayons of constant orientation.

these coefficients has been estimated by use of the familiar "Student" *t*-function. Significant positive correlation coefficients are found for the pairs of variables wet tenacity-orientation, wet elongation-orientation, and wet tenacity-wet elongation. These are expected from the correlation plots which were illustrated in Figs. 11–13. An especially interesting feature of the entire group of coefficients, however, is the fact that they indicate that lateral order as well as orientation has a very important effect on the physical properties of these yarns. Both wet tenacity and wet elongation are found to decrease significantly with increasing lateral order.

In each correlation, however, the specific relations sought are obscured by the fact that other variables have not been held constant. For example, both wet tenacity and orientation are known to be correlated with wet elongation and the observed coefficient for wet elongation with lateral order has undoubtedly been affected by the accompanying changes in wet tenacity and orientation.

In order to isolate the specific effect of a given structural parameter on a given property it is necessary to hold the other variable constant. This can be done mathematically by the methods of partial correlation which have been worked out by Ezekiel and others.[20] Linear relationships have been assumed between all six pairs of variables here since this greatly simplifies the calculation and since more complicated procedures have not so far appeared justified because of the scatter of the data.

When orientation and wet tenacity are maintained constant by the method of partial linear correlation, the relation between wet elongation and lateral order remains statistically significant, Table II. However, at constant wet elongation and orientation, wet tenacity is no longer significantly associated with lateral order, indicating that the apparent relationship shown by the total coefficient was merely the result of simultaneous variations in wet elongation and the dependence of wet tenacity upon wet elongation. The partial correlation coefficients also show that wet tenacity is correlated with orientation even at constant wet elongation and lateral

order, as might be expected. Wet tenacity is still significantly related to wet elongation even when orientation and lateral order are maintained constant, indicating that variables other than orientation and lateral order are important here. This correlation again suggests the similarity of stress-strain curve previously discussed and raises the question of what factors cause these yarns to fail at different elongations. Useful information concerning this problem might be obtained by studying the entire stress-strain curves rather than merely the properties at break. It is possible that the structure of some samples will correlate more closely with the elastic modulus than with tenacity and elongation at break.

When wet tenacity and lateral order for these samples are maintained constant, wet elongation is not significantly related to orientation. This indicates that the relationship observed between these two variables when wet tenacity and lateral order vary is merely the result of the close association between wet tenacity and wet elongation.

The Elongation-Lateral Order Relationship

The wet elongation/lateral order relation can be demonstrated graphically as well as by the analytical procedure just described. Orientation was maintained constant in Fig. 14 by merely plotting the data for all yarns having essentially the same orientation. The group was taken from the lower orientation end of the samples, where the differences in elongation tended to be larger. The close correlation between lateral order and wet elongation at constant orientation is clearly shown. Although lateral order as revealed by the radial intensity ratio has been used here, essentially the same correlation is observed if reciprocal interference widths are used, Fig. 15. A similar relationship between dry elongation and lateral order was also found, although the data tend to scatter more. Reciprocal interference widths have sometimes been used to calculate crystallite dimensions for native cellulose.[21] As mentioned previously, however, there are considerable uncertainties which make this specific interpretation of the x-ray data for a material such as regenerated cellulose appear of doubtful value.

Correction for Native Cellulose Content

Some of the changes in spinning conditions in producing the constant orientation samples just discussed also caused large changes in native cellulose structure content. The appearance of the native cellulose interferences in a typical pattern has already been illustrated in Fig. 5. As mentioned previously, the native cellulose can interfere seriously with determination of the lateral order parameters. In order to get some measure of the native cellulose content, the intensity ratio,

$$\left[\frac{I_{1n}-I_m}{I_1+I_{1n}-2I_m}\right]\times 100,$$

was calculated. I_{1n} is the intensity of the 101 native cellulose interference, I_1 is the 101 interference intensity for mercerized cellulose and I_m is the minimum intensity between the 101 and $10\bar{1}$ mercerized cellulose interferences, Fig. 5. Such a calculation is obviously crude because of the proximity of these interferences and their diffuse nature. However, lacking a better method, this procedure serves to grade samples according to apparent native cellulose interference intensity. The elongation-lateral order correlation can then be calculated just for the samples whose native cellulose intensity ratios fall within certain limits. An alternative correction method consists in adding a term for the native cellulose intensity $(I_{1n}-I_m)$ to the numerator of the radial intensity ratio. With either correction procedure, however, the correlation between wet elongation and lateral order is still statistically significant.

Applicability of the Relationships

These quantitative measurements adequately confirm the long suspected influence of disordered regions in cellulose, proposed by Mark[16] and Hermans,[15] as previously described. Baker, Fuller, and Pape have shown that toughness and extensibility of cellulose esters are favored by local chain disorder.[12] Baker points out that these phenomena are presumably general for all linear polymers,[10] a conclusion which is strongly

[21] H. Mark and K. H. Meyer, Zeits. f. physik. Chemie **B2**, 115 (1929).

supported by the present work on cellulose. It is emphasized, however, that these results do not mean that elongation of cellulose yarns will always increase as lateral order decreases. On the contrary, it would be expected that at constant orientation and lateral order other factors, which were nearly constant here, could operate to change the properties. There is some experimental evidence to be mentioned later that the properties of a yarn having the same orientation and lateral order as revealed by x-rays may vary according to the route followed in preparation. While some of these differences in properties may be caused by extraneous factors such as mechanical differences, it seems likely that structure also has had an important effect. For example, structural details not differentiated by x-ray diffraction, such as imperfections, dimensions, or amount of ordered regions, may have altered properties. The range of applicability and the limitations of this correlation between elongation and lateral order should become more clearly defined when information on structure and properties is available for yarns prepared under a wider variety of conditions.

Structural Changes During Spinning

Study of the yarn during each phase of preparation should be helpful in understanding the final structure and in accounting for properties.

Effect of Method of Coagulation and Regeneration

The timing of the coagulation and regeneration processes during spinning has an important effect on the lateral order and orientation of the final sample, as shown in Fig. 16. A yarn coagulated, stretched as the xanthate, then regenerated had much higher lateral order at the same orientation and could be oriented much more than a yarn which was simultaneously coagulated and partly regenerated before stretching. Regeneration occurring along with coagulation appears to fix a portion of the solution type disorder in the coagulating, shrinking structure. The regeneration occurring during coagulation allows the formation of interchain bonds which may restrict further ordering. Coagulation followed by regeneration evidently provides much more opportunity for ordering before the restricting cellulose

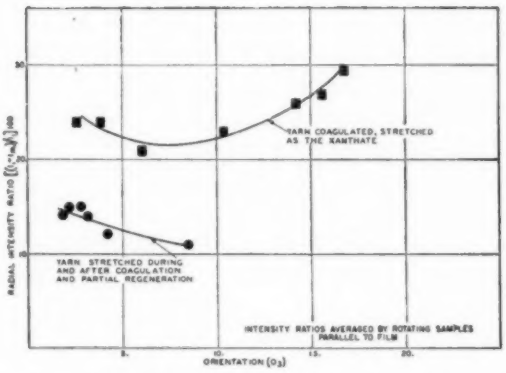

YARN COAGULATED, STRETCHED AS THE XANTHATE
YARN STRETCHED DURING AND AFTER COAGULATION AND PARTIAL REGENERATION
INTENSITY RATIOS AVERAGED BY ROTATING SAMPLES PARALLEL TO FILM
ORIENTATION (θ₂)

FIG. 16. Lateral order *vs.* orientation for yarns coagulated and regenerated by various methods.

interchain forces develop. It will be noticed that lateral order parameters for both yarns tend to decrease with increasing orientation in the low orientation range. The yarn stretched as xanthate could be oriented much more and at the higher orientation the order parameters pass through a minimum and finally increase with orientation. It is emphasized that orientation has an important effect on these relative lateral order parameters and it is not possible to decide from these data just how lateral order has changed with orientation; the data are closely comparable only at constant orientation. The lateral order parameters were calculated from patterns of rotated samples but this does not completely correct for differences in uniaxial orientation. The discrepancy is most important at the highest orientation and would make the order parameters too large. Both yarns show considerable selective orientation of the 101 planes at the lower orientations and this would also tend to increase the intensity ratio used as a measure of lateral order.

Skin-Core Structure

Study of some of these yarns is further complicated by a structural discontinuity across the filaments, the so-called skin effect described by Preston, Sisson, and others.[22] These investigators

[22] J. M. Preston, J. Soc. Chem. Ind. 199–203 (June 12, 1931); K. Ohara and R. Mizuno, Artificial Silk Stable Fiber J. Japan **1**, 10–12 (1933); K. Ohara, Sci. Pap. Inst. Phys. Chem. Research (Tokyo) **25**, 152–8 (1934); W. Schramek and J. Helm, Monatschr. Testil-Ind. **52**, 317–28 (1937); W. Schramek, Papier-Fabrikant **36**, 547 (1938); W. A. Sisson, talk at Gibson Island Textile Conference, 1943.

have reported that the proportion of skin can be varied widely by suitable changes in spinning conditions. In all such samples examined in this laboratory, an increase in skin thickness is accompanied by an increase in uniaxial orientation, a decrease in selective uniplanar orientation or preferential orientation of the 101 planes, and a decrease in lateral order. This evidence is only indirect but it suggests that the skin is more highly oriented uniaxially and has lower lateral order, while the core has a higher selective uniplanar orientation and lateral order.

Available data support the theory that the skin represents the proportion of the filament which was being coagulated and partly regenerated during the stretch in the region of the spinneret. The core is apparently still fluid viscose during the stretch near the spinneret and is coagulated and regenerated after the skin has set up and has reduced stretchability.

Deswelling and Relaxing

The deswelling or drying process in spinning has a large effect on fine structure. The disordering effect of drying from solvents less polar than water has already been discussed. Water wet cellulose yarns show considerable disordering of the 101 or hydroxyl-rich planes. The effect is observed both in gel cellulose and in rewetted yarn and the measured increase in 101 interference diffuseness appears to be larger than can be accounted for on the basis of mere superimposition of water scattering on the cellulose pattern.

The gel or freshly spun water-swollen yarn can be made either to decrease or increase in orientation on drying depending on the tension applied. Drying under low tensions or in a relaxed condition develops higher lateral order but lower orientation in the yarn than does drying under high tension. Similar changes in lateral order and orientation occur during heating a previously dried yarn in water at 90°C, as noted by Meyer and Mark.[23] The change in lateral order takes place during the initial heating in water, either relaxed or under tension. Subsequent wet heating changes orientation slightly, but has no effect on lateral order. Some very highly stretched yarns show a definite increase in lateral order merely upon wetting and drying relaxed. The changes in physical properties accompanying these changes in lateral order vary depending upon the particular treatment. However, after either drying relaxed or heating in water relaxed the elongation-to-break is increased relative to that of a suitable control. Since both orientations and lateral order change, it is difficult to say just how these latter after-treatment effects fit in with the property-structure correlations already discussed for yarns spun under various conditions. However, as mentioned, it can be said that factors other than orientation and lateral order also have important effects.

The lateral order of freshly spun yarns is therefore metastable in that it is measurably increased by relaxing or heating in water, just as the celluloses of low order obtained by regenerating or deswelling in non-aqueous media were ordered by wetting out at room temperature. In general, however, the disordered structures obtained by a particular mode of spinning have appeared to be more stable than those obtained by non-aqueous regeneration or deswelling. These considerations suggest that lateral disordering introduced after the yarn structure is set up is of a more localized nature than that existing in the original structure.

ACKNOWLEDGMENT

The author wishes to acknowledge with appreciation the helpful suggestions and valuable samples, apparatus, and techniques obtained from Dr. O. E. Schupp and Dr. C. W. J. Wende of the Chemical Department and from members of the Rayon Technical Division, E. I. du Pont de Nemours & Company. Especial thanks are expressed to Dr. N. L. Cox of the Chemical Department for preparing all of the experimental viscose rayons described here, to Dr. A. F. Smith of the Chemical Department for continued guidance and interest in this work, and to the Rayon Technical Division, E. I. du Pont de Nemours & Company, for generous support of this research.

[23] K. H. Meyer and H. Mark, *Der Aufbau Der Hochpolymeren Organischen Naturstoffe* (Akad. Verlagsgesellschaft M. B. H., Leipzig, 1930), first edition, p. 183.

A Theoretical Comparison of the Visual, Aural, and Meter Reception of Pulsed Signals in the Presence of Noise*

J. H. Van Vleck and David Middleton
Research Laboratory of Physics, Harvard University, Cambridge, Massachusetts
(Received July 9, 1946)

The present paper examines theoretically the relative sensitivity of the detection of signal pulses in the presence of noise, (a) by observation of an oscilloscope, (b) by aural perception, in which one listens to the fundamental or a low harmonic of the pulse repetition frequency (PRF), (c) and by a meter. The metering scheme may be either aperiodic, where the rectified current is fed directly to a meter with a long time-constant, or periodic, where the rectified current is sent through an audio-filter tuned to the PRF, given a supplementary rectification, and then passed through the meter. The dependence of the sensitivities of the different methods on various relevant parameters is studied in some detail. These parameters include the width and the shape of the IF response, the pulse length, the PRF, and in aural or meter reception, the duration of the gate, the width of the audio-filter, and/or the time constant of the meter. The descriptive survey of the results is given in Part I and the mathematical analysis in Part II. Among the more important results are (I): The optimum IF filter is the conjugate of the Fourier transform of the pulse, not merely for visual reception, as was previously known, but also for aural or meter reception as well. (II): For very weak signals the linear detector requires only about 5 percent more input signal power than does the quadratic to achieve the same minimum detectable signal (same final signal-to-noise ratio). (III): The aperiodic meter has the advantage of not requiring knowledge of the PRF, and has potentially great sensitivity if spurious fluctuations in gain can be balanced out. (IV): Meter methods can be made more sensitive than the oscilloscope if long time-constants are available. Gating is also necessary. (V): Although the best IF filter is the Fourier transform of the pulse, the best pulse is not the Fourier transform of the filter in aural reception (though it is in visual), for the best results in meter or audio-detection are obtained by using long pulses. In visual work, the pulse length is immaterial, to a first approximation. Curves are given showing the power required to achieve a given signal-to-noise ratio as a function of pulse length, IF filter width, the PRF, gating time, and audio-filter width, in some cases when the pulse and filter are not matched (i.e., are not related as Fourier transforms). Some numerical estimates of aural and meter performance relative to visual are also essayed.

PART I. DISCUSSION

I. Introduction and Formulation of the Problem

THE purpose of the present paper is a theoretical comparison of the merits of visual (oscilloscope) and aural detection of pulsed signals, with particular emphasis on the best intermediate frequency (IF) filter design for the two types of reception. It is not *a priori* obvious that the optimum design is the same in the two cases. By visual reception we mean that the modulation is spread out on a screen by means of a cathode-ray oscilloscope. By aural we imply that the emphasis is on detecting a particular tone (the fundamental of the pulse repetition frequency (PRF), or a low harmonic thereof) in the audio-region. Here at the outset we should emphasize that in the present paper we usually employ the term *aural reception* to include any detection device which is actuated by one particular harmonic component of the PRF, or in other words, involves integration in time rather than the observation of the instantaneous effect of the superposition of all harmonic components, as in visual perception. The instrument need not be the human ear, but instead can be, for instance, a meter, or recording device. The harmonic component, furthermore, under certain circumstances, can be the constant or d.c. term, not involving the pulse repetition period, as in so-called *aperiodic* metering schemes.

In the visual case, the best filter, in the first approximation, is that which gives a maximum value of the ratio of the peak signal to the noise background. This statement can be formulated more precisely as follows: coming into the IF filter there is a fluctuating noise voltage and also a signal $S_I(t)$, which is a definite function of time. After filtering and rectification, the video amplitude is not expressible as the simple sum of pure signal and pure (i.e., normal, random) noise, as there are complicated modulation or interference effects between the two, especially

* This paper is based in whole on work done for the OSRD under contract OEMsr-411 with the President and Fellows of Harvard College.

for the linear detector. If the signal is periodically repeated, there will be a particular epoch at which the amplitude on the oscilloscope screen is on the average a maximum. *The excess of the average of the amplitude at this epoch over the mean amplitude in the absence of signal we denote as* s_v. We say on the average, because sometimes the noise may interfere constructively and sometimes destructively with the signal. We must seek, therefore, to make $s_v(t)$ stand out as much as possible relative to the r.m.s. deflection n_v caused by the noise background. The latter may be computed on the assumption that the signal is zero, since the duty cycle (the "on"-period of the pulse), is presumably very small, and hence the cross-modulation terms contribute little. (See the paragraph following Eq. (16), Part II.) According to this viewpoint the best filter is that which maximizes s_v/n_v. This we shall refer to throughout the paper as the "simple criterion" for visual detection and is that which we shall use as our standard. We employ the adjective "simple" because actually examination merely of the ratio s_v/n_v neglects certain factors. It does not consider the effect of the width rather than the height of a "pip" and of the "blades of grass," and especially it does not adequately take into account the effect of pulse repetition. The more frequent the pulse, the more easily the eye integrates, in the limited time at its disposal, to detect the mean position of the pip relative to a fluctuating background. Integration problems in visual detection are beyond the scope of this work, as they have been considered elsewhere, especially by North,[1] Uhlenbeck,[2] and others.[3,4] In discussing the results we shall mention the modifications introduced by the more refined criteria.

With audio-reception the point is to make the ratio of energy in the fundamental of the PRF as large as possible relative to the noise in a small frequency interval immediately surrounding the discrete frequency equal to the PRF. The audio-reception may be accomplished in either of two ways. In case the PRF is known, we may insert a filter a few cycles wide immediately surrounding the PRF, so as to cut down the noise. The presence of the signal can then be detected not necessarily by the ear but in principle, say, by lighting a light—more practically, by inserting a thermocouple, or better still, by means of an additional rectification in connection with a meter, as will be discussed more thoroughly in Section IV, Part I. The sensitivity of the device is contingent on the ability to use a very narrow filter, since the signal energy is concentrated at discrete frequencies, whereas noise is spread through a continuum.

In case the PRF is not known, one does not know where to put the filter. The apparatus based on light indication or metering (unless of the aperiodic variety) then obviously becomes quite unworkable in general. There is, however, one instrument which is able to construct a moderately narrow filter about the PRF without knowing where the latter is. This instrument is the human ear, which has the remarkable property that it can pick out energy concentrated in a particular periodic component even though the latter is greatly over-shadowed in total amount by the background energy in the continuum. Namely, the ability of the ear to hear a unifrequentic or pure tone f_p is masked only by the noise background in a particular frequency interval $f_p-\Delta f_e/2$ to $f_p+\Delta f_e/2$, centered about f_p, so that effectively the ear acts like a filter of width Δf_e. Of course, the filtering action of the ear is not perfectly rectangular as we assume it, but the error introduced by this idealization is not serious for our purposes. We will not attempt to go into the physiology of why

[1] D. O. North, in an unpublished report PTR-6C, entitled "Analysis of factors which determine signal-noise discrimination in pulsed carrier systems." (RCA, Princeton, June 25, 1943.)

[2] G. E. Uhlenbeck and M. C. Wang have carried through similar studies which it is hoped will be available soon at the Massachusetts Institute of Technology's Radiation Laboratory. In this connection, very extensive experimental work has been done on all phases of the visual problem by J. W. Lawson, also at (M.I.T.) Radiation Laboratory. It is expected that this work will be published by 1947, appearing in the appropriate Handbook. (Book 24, Radiation Laboratory Series, "Threshold Signals," by Lawson and Uhlenbeck.)

[3] P. J. Sutro, "Theoretical effect of integration on visibility of weak signals through noise." (Harvard Radio Research Laboratory, Report 411–77, February 14, 1944.) Available at Harvard War Archives, Littauer Building, Harvard University.

[4] E. R. Brill, "Improvement in minimum detectable signal in noise through the use of the long afterglow tube and through photographic integration." (Harvard Radio Research Laboratory, Report 411–84, February 8, 1944.) Available at Harvard War Archives, Littauer Building, Harvard University.

the ear has this behavior, but it is probably due to the fact that it (the ear) embodies a collection of nerves, each responding to a very narrow frequency band, and is capable of discriminating when one nerve is more strongly excited than the average. The effective band width is a function of f_p, the dependence being shown in Fig. 1, taken from data in a paper by Harvey Fletcher.[5] For the values of f_p in which we are interested, the quantity Δf_e is of the order 50 to 200 cycles. The signals with which one is concerned involve many harmonics as well as the fundamental, but Fig. 1 shows that the effective filter width is greater for higher frequencies, so that they are usually more readily screened by the noise background; hence it is better not to use harmonics, and so we gauge audibility by the masking of the fundamental. It is important to note, however, that when the PRF is much below 1000 c.p.s. it may be preferable to use a suitable harmonic of the fundamental, since the ear is most sensitive to signals in the range 1–2 kilocycles.

According to the above, then, any filter centered around f_p and wider than the corresponding Δf_e is of no particular help, while any filter of width $\Delta f'_e$ narrower than Δf_e has the effect of reducing the power level of the noise by a factor $\Delta f'_e/\Delta f_e$. The point which we wish to emphasize is that *the mathematical criterion for effective reception by the ear is the ratio of signal energy in a discrete Fourier component to the energy in the immediate vicinity of this component*, the "immediate vicinity" being the natural selectivity of the ear or the width of the artificial filter, whichever is narrower. The use of the latter, of course, is possible only if the PRF is known.

In either visual or aural reception there is a certain critical threshold of signal to noise which must be exceeded before the signal is perceived. However, from the above we see that the nature, and so presumably the magnitude, of the threshold is different in the two cases. Namely, for the visual case the critical ratio was that of the average maximum signal to average total video noise amplitude s_v/n_v. On the other hand,

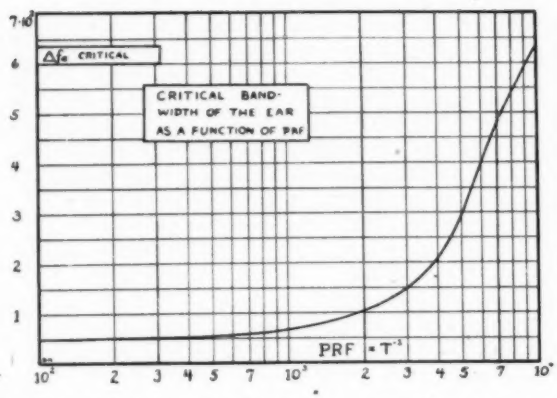


FIG. 1. Critical band width of the ear as a function of the pulse repetition frequency.

in aural detection the critical ratio s_a/n_a is that of one harmonic signal component to the r.m.s. amplitude of the portion of the noise which is confined to the spectral region covered by the critical masking zone of the ear (Fig. 1) or by the audio-filter, if narrower.

The question of the numerical results of the minimum detectable values of $(s/n)_v$ and $(s/n)_a$ obviously depends on the observer. In oscilloscope measurements the minimum detectable signal-to-noise ratio is usually expressed by the experimentalists in terms of the amplitude ratio s_F/n_F where s_F and n_F are, respectively, the mean amplitudes of signal and noise after filtering by the IF, but before rectification, rather than in terms of s_v/n_v directly. Very acute observers can detect values of s_F/n_F as low as $10^{-\frac{1}{2}}$, i.e., -10 db, whereas an unskilled person may only see the signal when s_F/s_F is almost unity, which is probably the safer estimate to use under conditions where there are many distractions. (These figures are based on essentially infinite integration time, which for most people is practically of the order of one minute, sometimes running as high as about five minutes under the most favorable conditions before further improvement is unnoticeable.) In the aural case, Fletcher[5] gives the threshold s_a/n_a as of the order of magnitude unity. Some recent rather crude experiments undertaken by the authors indicate that to perceive a value $s_a/n_a=1$, the observer must be existing under rather ideal laboratory conditions, and that with any distraction, mental or otherwise, the minimum detectable value of $(s/n)_a$ is more nearly of the order $10^{\frac{1}{2}}$ (+10.0 db). These

[5] H. Fletcher, Bell Lab. Monograph B-1205 (Dec. 1938), cf. Fig. 17 of "Auditory Patterns as a Means of Studying the Hearing Process Taking Place when Sound is Sensed."

two values, 0 db and +10.0 db (or correspondingly, −10.0 db and 0 db in the visual case), we shall refer to as the "optimist's" and "pessimist's" estimates, respectively.

The criteria of the different minimum detectable signals for visual and aural reception may now be specified in analytic form. Throughout the paper we are unconcerned with the first rectification (or mixing) as this is essentially a linear process which has the effect of displacing the carrier frequency from an RF value f_{rf} to an IF value f_0. After the IF filter the mixture of signal and noise is sent through a rectifier, here taken alternatively as quadratic (characteristic, $f(x)=x^2$), or linear, without bias (with the characteristic $f(x)=x$, $x>0$, $f(x)=0$, $x<0$). The signal disturbance is denoted by $s_v(t)$. In visual detection the minimum observable signal-to-noise ratio is some particular value of $s_v(t)/n_v$, where n_v^2 is the total video noise power, given by

$$n_v^2=\int_0^a W_{LF}(f)\frac{d\omega}{2\pi},$$

where a is a limit great compared to the IF filter breadth but small compared to $2\pi f_0$, and where $W_{LF}(f)$ is the spectral density of the mean square noise amplitude. (See Eqs. (18), (26), Part II.)

In audio-detection, on the other hand, one is concerned with the energy in a given harmonic component, divided by the energy in the immediate region surrounding this component. Therefore the effective signal-to-noise ratio is

$$s_a/n_a=\frac{2^{\frac{1}{2}}}{T}\left|\int_{-T/2}^{T/2}s_v(t)\exp(ik\omega_p t)dt\right|\Big/[W_{LF}(kf_p)\Delta f_e]^{\frac{1}{2}},$$

where k is the order of the harmonic component kf_p being received, $\omega_p=2\pi/T$ is the pulse repetition (angular) frequency, and Δf_e is the filter width. (The filter is regarded as rectangular; if it is not, an equivalent rectangular width which admits the same amount of power for a filter which has the same maximum response is to be used.)

Throughout the paper certain simplifications are made in the expression for s_a/n_a. In the first place, it is assumed that the pulses do not overlap appreciably, which is certainly true if the width of the IF filter is large compared to the PRF. The limits of integration in taking the Fourier component can then be replaced by $+\infty$ and $-\infty$, appropriately, if $s_v(t)$ is then understood to be only the disturbance due to one pulse. Further, the length (in time) of one pulse is supposed short compared to the pulse repetition period, so that the factor $\exp(ik\omega_p t)$ may be taken as nearly constant in the region where $s_v(t)$ is appreciably different from zero. This is true unless the harmonic is so high as to be comparable with the reciprocal of the pulse length. Since the PRF is small compared to the noise band width, we may assume that the noise background $W_{LF}(k\omega_p)$ has substantially the same value as in the vicinity of zero frequency, making the above expression for s_a/n_a take the form

$$s_a/n_a=\frac{2^{\frac{1}{2}}}{T}\int_{-\infty}^{\infty}s_v(t)dt\Big/[W_{LF}(0)\Delta f_e]^{\frac{1}{2}}.$$

This expression is independent of k ($k=1, 2, \cdots$), so that it is immaterial which harmonic component is utilized.* The input power required to achieve a given critical value of the ratio $(s/n)_v$ or $(s/n)_a$ will be a function of the following variables:

(a) the width and shape of the IF,
(b) the width and shape of the pulse,
(c) the pulse repetition frequency (PRF),
(d) the duration of the gate,
(e) whether the rectifier is linear or quadratic,
(f) whether the detection is visual, aural, or by meter, and
(g) the width of the audio- (or meter) filter.

It is the purpose of this paper to examine how the input power depends on any one of the above variables, when the other six are held fast.

II. Best IF Filter Width

We begin by studying the optimum design of the IF filter. Here there is a very simple result. Namely, the best IF filter is one which is the conjugate of the Fourier transform of the pulse

* Here the choice of harmonics is unimportant from the point of view of the *intensity* level of the signal but not from the *loudness* level, which is a measure of the actual perception of the ear, and which varies with frequency. The intensity level merely measures the sound energy per unit area falling on an intensity-level meter, and is not the subjective intensity, or "loudness" of the perceived sound in the ear.

shape (except possibly for a phase shift which varies linearly with frequency). That is to say, if the modulated pulse presented to the IF is $\sqrt{2}s_I(t)\cos\omega_0 t$, and if the Fourier transform of $s_I(t)/\sqrt{2}$ is denoted by

$$S_I(\omega)=\frac{1}{\sqrt{2}}\int_{-\infty}^{\infty}s_I(t)e^{-i\omega t}dt,$$

so that

$$s_I(t)=\frac{1}{\pi\sqrt{2}}\int_{-\infty}^{\infty}S_I(\omega)e^{i\omega t}d\omega,$$

then the best filter design is one whose amplitude response is

$$G_{IF}(\omega)_{OPT}=\alpha S_I(\omega)^*\exp(-i\omega t_0).$$

Here α is a trivial constant factor calibrating the absolute sensitivity of the filter. The origin of frequency is taken at the carrier, so that the point of maximum response is $\omega=0$. Usually the pulse is symmetrical about its maximum, and then G_{IF} is real, the distinction between $G_{IF}(\omega)$ and its conjugate $G_{IF}(\omega)^*$ vanishing. This result is demonstrated in Part II, Sec. III, cf. Eq. (53).

The criterion is general. It applies regardless of whether one has audio- or video-reception, or whether one has a linear or square-law detector. The existence of the Fourier transform criterion has already been noted, in the visual case by North[1] and others. In the instance of aural reception, the phase of $G_{IF}(\omega)_{OPT}$ is immaterial, and it is only necessary that $|G_{IF}(\omega)_{OPT}| \sim |S_I(\omega)|$ (*vide* Eq. (56)).

The phase shift constant t_0 is arbitrary as far as the optimum quality of the filter is concerned, i.e., it does not affect its performance. A change in the value of t_0 merely gives a different time delay of a pulse transmitted by the filter. Actually, any physical filter will have a definite value of t_0, as discussed in more detail in Sec. III of Part II, but knowledge of t_0 has no bearing on the discussion of filter efficiency.

It is particularly interesting that the Fourier transform criterion applies even if the reception is by audio means. The aural case does not appear to have been too clearly understood. Some, for instance, have thought that it would be best to have a very narrow filter (much narrower than the Fourier transform of the pulse) just because in aural reception one is immediately concerned with only a narrow band of frequencies. However, this is not correct—the best IF filter will be just as wide as for video reception. See Sec. III (b), Part II, for the details of the demonstration.

A filter whose characteristic satisfies the optimum condition, i.e., is the conjugate of the Fourier transform of the pulse, we shall henceforth refer to as a *matched* filter. It has sometimes been suggested that in visual reception results will be improved by adding a video-filter, i.e., a filter following rectification. If, however, the IF filter is already matched; the insertion of such a video-filter theoretically gives no improvement. (See Book 24, M.I.T. Radiation Laboratory Series, when it appears.)

III. The Linear *vs.* the Quadratic Detector

On the whole the results for the half-wave linear and quadratic rectifiers are remarkably similar either in aural or visual reception. This is particularly true if one has sensitive apparatus and/or good observers, so that the values of the signal-to-noise ratio s_F/n_F are reasonably small compared to unity. Then the behavior of the linear and quadratic devices is the same as regards the value of s_v/n_v, corresponding to a given pulse system and IF filter, except for a numerical factor. The latter proves to have a value $0.957=\frac{1}{2}(4/\pi-1)^{-\frac{1}{2}}$, as has already been noted by North[1] in his interesting paper. In the audio case the constant is even more nearly unity, being somewhat dependent on pulse shape. For the specific example of gaussian pulses and filters, considered analytically in Secs. IV and VI of Part II, the constant is 0.973. The result in the video cases is contingent only on small s_F/n_F and does not demand a particular filter shape, (e.g., gaussian), or matching of the IF filter to the pulse. The mathematical reason for this similarity in the video case is that for small values of s_F/n_F the signal amplitude transmitted by the rectifier is given by a monotonically increasing series in the ratio s_F/n_F of the form $0.957\{(s_F/n_F)^2+\cdots\}$. If the first term only is kept, a quadratic behavior follows, like that of the square detector whose first term is $(s_F/n_F)^2$, Eq. (46). The absence of the linear term s_F/n_F is the analytical embodiment of what the engineer

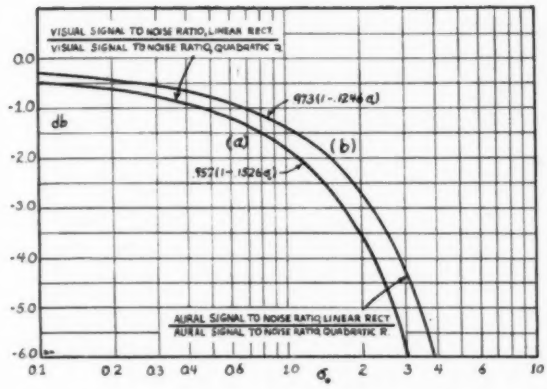


FIG. 2. A comparison of the linear and quadratic detectors in aural and visual reception at match. Curve (a) is a comparison of the linear with the quadratic detector for visual reception, while curve (b) is the same comparison in aural reception, a gaussian pulse being assumed. Here σ_0 is the ratio of the peak signal to r.m.s. noise amplitude leaving the IF.

calls *modulation suppression*, and makes the transmitted signal very much less than it would be otherwise. (See Sec. II (d), Part II.)

The retention of only this leading term, however, always overestimates the performance of the linear detector. A more accurate expression for not too large values of s_F/n_F requires that the inferiority factor for the linear rectifier in video reception be $0.957(1-0.1526 s_F/n_F)$ instead of 0.957. This is a sort of interpolation formula valid up to about $s_F/n_F = 10^{\frac{1}{2}}$. For large values of s_F/n_F the signal amplitude becomes obviously linear in s_F, and the noise rather than the signal is suppressed. The second term of the interpolation formula represents, so to speak, a correction for the tendency of the instrument's response to become ultimately linear rather than quadratic. The relative performance of the two types of detector in the video case is shown in curve (a) of Fig. 2, which is obtained by dividing Eq. (47) by (46) (see also Eq. (82)).

In the aural case a general expression for the relative efficiency of the linear and quadratic detectors cannot be given except for small s_F/n_F until the shape of the pulse and filter are specialized. One can, however, say that when s_F/n_F is not appreciable, i.e., $s_F/n_F < 1.0$ the inferiority factor of the linear detector is always less than about 1.5 db, and the deviations from quadratic performance are of the same order of magnitude as in the video case. This is discussed at somewhat greater length in Part II, Secs. II (e), and Sec. IV (c). Curve (b), Fig. 2 is obtained from Eq. (84) and illustrates the relative aural performance of the detectors, when the incoming pulse is gaussian.

IV. Use of a Meter

It is particularly to be noted that with aural reception the background noise power n_a^2 is directly proportional to the width of the audio-filter. Hence, in principle, if one can construct arbitrarily narrow filters and if the signal is exactly periodic so that its Fourier component is strictly a "delta"-function (i.e., a function with an infinitely high and narrow peak at some point $x=x_0$, and zero everywhere else, $x \neq x_0$) one should, for a given minimum detectable value of $(s/n)_a$, be able to achieve infinitely high sensitivity in terms of s_F/n_F, simply by narrowing the audio-filter indefinitely. However, there are of course, practical limitations as to how much a filter can be narrowed, *viz.* (a), purely constructional difficulties, (b) insufficient stabilization of the PRF, and (c) limitations because the targets are not in the path of the beam an infinitely long time, so that we have a signal repeated a finite rather than an infinite number of times. The signal spectrum consequently is not unifrequentic, but is "smeared out," having Fourier components which are distributed over a finite interval instead of being sharp delta-functions. Probably (b) is not too serious a difficulty. As regards (a) it is not easy to build audio-filters even as narrow as 10 cycles, and still narrower filters would require lock-in amplifiers, which are fairly clumsy and bulky pieces of apparatus. Effect (c) is always present to some extent.

Aperiodic Methods

Although a narrow filter is difficult to construct in the audio-region, on the other hand, a very narrow low pass filter, which includes d.c., is easy to obtain simply by sending the rectified current through a condenser in parallel with a meter, which is used as the recording instrument. Consequently, the idea may occur to detect the constant term of the Fourier expansion of the signal, rather than the fundamental. As the constant term does not involve the PRF in any way,

instruments based on the recording of this term may be called "aperiodic."

The constant term in the Fourier expansion has associated with it an energy half as great as any other component.* This reduction by a factor $\frac{1}{2}$ is offset by the fact that the low pass filter extends on only one side of the origin in frequency, so that the noise continuum transmitted is halved. Hence our various expressions for signal-to-noise ratio, etc., in aural detection also apply to the aperiodic metering scheme if we use in the formulas an effective filter width equal to ***double*** the width of the physical low pass filter.

As far as the competition between the signal and the fluctuations in the continuous noise background is concerned, the zero frequency (also called aperiodic) method is thus satisfactory. The difficulty comes from masking due to spurious variations in the d.c. due to noise, or in other words, from what the engineer calls "changes in gain." When we say spurious, we mean variations caused by factors (e.g., temperature, apparatus changes, etc.) other than statistical fluctuations like those in the Shot effect, etc. From the purely statistical standpoint, the d.c. due to noise is constant in time, and can, in principle, be balanced out by a proper bias. However, the d.c. due to the noise is enormously larger than the continuous noise background admitted by the filter, and hence much greater than the signal, as the latter two quantities are comparable. Namely, the energy in the d.c. exactly equals that in the ***whole*** low frequency noise continuum, for a square detector and exceeds it by a factor $3.7=(4/\pi-1)^{-1}$ for the linear rectifier. Since the width of the low frequency continuum is of the order of the IF frequency width we see that the d.c. noise will exceed the fluctuating noise continuum admitted by the filter by a factor of the order of the ratio of the IF filter width to that of the audio. This ratio is about 10^5–10^6. Hence an infinitesimal spurious variation in the d.c. noise voltage will overshadow and obscure the modulation due to the signal, and so make the device not at all feasible, unless it has very special equipment for balancing out spurious fluctuations.

If the oscillations in gain are very slow, they can be avoided simply by inserting a blocking condenser. The filter is then effectively not of the low pass type, but instead cuts off at very low frequencies, so that pure d.c. is eliminated. This artifice will work only if the spurious changes in gain have a longer "time constant" than does the signal modulation, so that the filter will pass the latter but not the former. Unfortunately, the two times are rather likely to be comparable in many applications.

Periodic Methods

There is one way of escaping from this difficulty of gain fluctuations, and this is to use a meter in conjunction with an audio-filter. This procedure requires rectification. Current confined to the frequency range passed by the audio-filter (centered about, say, the fundamental of the PRF), is rectified, and enters an electromechanical low pass filter which admits only d.c. and extremely low frequencies (for example less than $\frac{1}{10}$ or 1 cycle as an upper limit). To avoid constructional difficulties, the audio-filter will not be nearly so narrow as the meter filter. The presence of the signal then shows up as a change in the amount of d.c. It may seem at first sight that this scheme is exactly like the aperiodic one just discussed, but the insertion of the additional rectification* makes a great difference. In other words, the action of

(A) rectifier→low pass (directly) is quite different from

(B) rectifier→audio-pass→rectifier→low pass.

Obviously (B) requires knowledge of the PRF, whereas (A) does not. Hence (A) and (B) may be called, respectively, the aperiodic and the periodic metering methods.

In (A) as in (B) after the final rectification, the energy in the d.c. due to noise is about equal to the energy in the continuum. The continuum, however, has a width comparable in (A) with

* The reason is that the constant term has an amplitude $\left|(1/T)\int_{-\infty}^{\infty} s_v(t)dt\right|$, rather than $\left|(2/T)\int_{-\infty}^{\infty} s_v(t)dt\right|$, but the reduction in intensity is by a factor $\frac{1}{2}$ instead of $(\frac{1}{2})^2$, because for the constant term there is no distinction between r.m.s. and peak amplitude.

* *Note:* The audio-filter is always sufficiently narrow that the noise entering the final rectifier may be considered "pure" and random, for if a small portion of the spectrum alone is taken, randomness is restored, and the gaussian characteristics of pure noise are manifest.

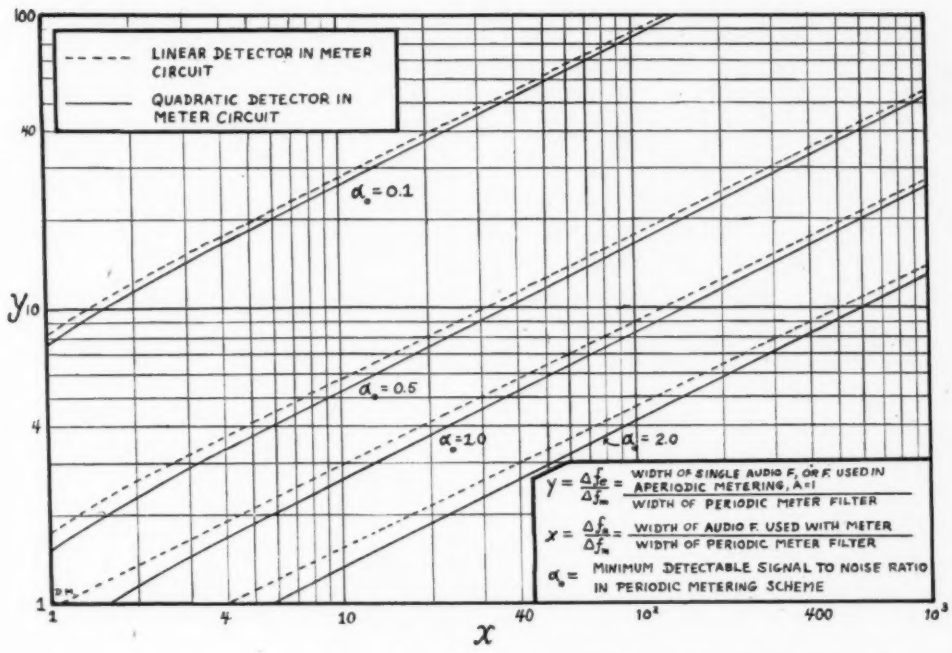

LINEAR DETECTOR IN METER CIRCUIT
QUADRATIC DETECTOR IN METER CIRCUIT
$\alpha_0 = 0.1$
$\alpha_0 = 0.5$
$\alpha_0 = 1.0$
$\alpha_0 = 2.0$
$y = \frac{\Delta f_e}{\Delta f_m} =$ WIDTH OF SINGLE AUDIO F, OR F USED IN APERIODIC METERING, A=1 / WIDTH OF PERIODIC METER FILTER
$x = \frac{\Delta f_a}{\Delta f_m} =$ WIDTH OF AUDIO F USED WITH METER / WIDTH OF PERIODIC METER FILTER
$\alpha_0 =$ MINIMUM DETECTABLE SIGNAL TO NOISE RATIO IN PERIODIC METERING SCHEME
y_{10}
x
100
40
4
1
1
4
10
40
10^2
400
10^3

FIG. 3. Width of the equivalent earphone filter Δf_e used in aural detection of the minimum perceivable signal in the presence of noise as a function of the audio-filter width Δf_a used in meter detection of the same signal-to-noise ratio α_0. The two widths compared are both normalized in the figure by division by the quantity Δf_m, which is twice the width of the final low pass filter in the periodic metering scheme. It is assumed that the minimum discernible signal-to-noise ratio in the final circuit is the same for aural or meter detection.

the width of the IF filter, and in (B) with that of the audio-filter. Hence the final continuum is much narrower in (B) than in (A). As a result, the low pass filter admits a very small fraction of the total continuum energy in (A) but an appreciable fraction of it in (B). Assuming that the minimum detectable signal (apart from troubles due to spurious variations in d.c.) corresponds to the low pass filter receiving a signal comparable with the noise background, we see that in (B) the signal will be comparable (say within a factor 10 or 100) with the d.c. of the noise, instead of being a million or so times smaller than the latter. Thus spurious variations in d.c. intensity will not be lethal. (We have here assumed that the discrete noise at the PRF in the audio-filter has been balanced out by some method such as the Jordan double-gate scheme, described later. If this is not so, the signal may be more nearly one-thousandth or one ten-thousandth of the noise d.c., but still the latter is not so overwhelming as in the straight aperiodic metering system.)

The mathematical theory of the extra rectification involved in the periodic meter is treated in Sections V and VI, Part II. The theory is in the main the same as that for the rectification following the IF, with the audio-filter now playing a role similar to that which the IF filter did previously. However, there is one simplification. The signal transmitted by the audio-filter is sinusoidal rather than pulsed, so that its amplitude can be treated as constant.

There are some false impressions to be guarded against in considering the periodic meter. In the first place, for the periodic system (B), a very narrow reception band in the final circuit (e.g., 0.05–0.5 cycle), with a narrow audio before the extra rectification (e.g., 5–10 cycles), does not represent nearly so great a sensitivity as would be obtained if one had such good filter construction that one could put an equivalent very narrow band (say, 0.05–0.5 cycle) directly in the audio-circuit. The reason is the non-linearity of the rectification. Namely, if the audio signal-to-noise ratio is small (and if it were not, there would be no need for the meter), the response of the rectifier is essentially quadratic in this

ratio, even if it is of the linear rather than the square variety, because of modulation suppression, as shown by Eq. (97).

Curves illustrating the effect discussed in the preceding paragraph are given in Fig. 3. They are drawn for a linear rectifier in the meter circuit (dotted line), but the results are not materially different for the quadratic case (solid line), especially for weak signals. These curves correspond to various assumed minimum detectable signal-to-noise ratios for the meter, and the abscissa is the ratio of the audio to meter filter width. The ordinate is the audio-filter width, measured in multiples of the meter filter width, which would give the same signal-to-noise ratio in the audio-circuit, without the meter, as obtains with the meter in the final circuit. In other words, the ratio of ordinate to abscissa measures the factor by which the audio-filter would have to be diminished in order to permit one to dispense with the meter circuit. For the typical case of a minimum detectable signal-to-noise of unity, a linear second detector, and a meter width of one cycle, $\Delta f_m = 2$, and with a ten-cycle audio-filter $\Delta f_a = 10$ the equivalent *single* audio-filter is 4.3 c.p.s. As an example of another typical case, assume again that the minimum detectable signal-to-noise ratio is unity, the width ($\Delta f_m = \frac{1}{10}$) of the meter filter is 1/20 cycle, and that of the audio is now 5 cycles. The equivalent single audio-filter is then 0.65 c.p.s. and the equivalent low pass filter is 0.33 c.p.s. These figures are based on the assumption that the filters are of gaussian shape, and the width is determined by the half-power points. In case all the filters are nearly square, or "ideal," the formulas still apply if one takes the mathematical quantity Δf to be 1.06 times the actual filter width.

Considerations Relative to Time-Constants of Meter

In either periodic or aperiodic methods the noise is counterbalanced by a suitable bias, and the *mean* meter reading can thus be made zero in the absence of signal. The presence of the signal will cause a systematic deflection of the needle, if the time-constant of the meter is small. Thus for a comparatively wide meter filter, the limiting factor preventing observation is the irregular fluctuation due to the noise continuum.*

The question of how much trouble is caused by the long time-constant of a meter circuit appears to be appraised very differently with a discouraging disparity of opinion. Some claim extraordinary sensitivity with meter reception, whereas others claim that the time-constant difficulties are so serious that, for practical purposes, a meter circuit is to all intents and purposes useless on an airplane or for other operations. Perhaps a conservative viewpoint is that a meter circuit will not give too much gain in efficiency unless the target is under observation in the receiver for at least a second and this, of course, requires that the angular scanning be rather slow. But in circumstances where long time-constants, say one to six or eight seconds or more, are allowable, the (periodic or aperiodic) meter, or its cousin, the recording tape, may be able to compete favorably with other methods of perception, for example, with aural detection where the strain of "listening through" is much greater than for "looking through." There seems to be insufficient experimental evidence at this time, however, as to precisely what a meter can and cannot do compared with that for other systems. So before a definite answer can be given, we must know at least what the order of magnitude of the minimum perceivable signal on the meter is. At any rate, one is certainly not *a priori* justified in discarding meters as detecting devices.

The experimental evidence available to us at present seems to indicate that meters with short time constants, i.e., wide filters, say, 1–2 cycles, are not as effective as other methods when the situation is such that the time of observation of a

---

* *Note:* One point should be noted in connection with these fluctuations. They are themselves modified by the presence of the signal, as the cross-terms between signal and noise make the continuum energy different from what it is with noise alone. This effect was wanting in previous parts of the analysis, as the pulsing of the signal made the signal-noise cross terms unimportant because they existed only a small fraction of the time. The meter, however, is exposed solely to a sinusoidal signal since the audio-filter excludes the harmonics which interfere with the fundamental to shape the pulses. We do not endeavor to calculate the modification of the continuum energy by the signal (which one of us hopes to treat in a later paper), because it is difficult, because experimentally it does not seem to be an effect of much importance, and because it is not clear whether the increase of the "wobble" (in distinction from systematic drift) of the indicator when the signal is on makes the latter easier or more difficult to perceive.

target si brief and the conditions difficult. However, in operations where longer intervals of observation are allowed, say 2–10 seconds, or longer, corresponding approximately to meter filter widths of $\frac{1}{2}-\frac{1}{10}$ cycles the meter may offer definite advantages over other systems of perception. For an example, consult the table, Sec. VII, Part I.

Clearly, the minimum detectable signal-to-noise ratio in a meter depends a great deal on tactical considerations and will be higher in the field than in the laboratory. In the table, Sec. VII, we have assumed that s_m/n_m is unity. This whole question of the minimum perceivable signal in meter detection, as well as in the other systems, devolves ultimately upon memory, whether it be psychological, i.e., remembering a series of successive readings, or mechanical, in which the memory process is achieved by means of a recording device, such as an Esterline-Angus recording meter, for example. Under ordinary conditions a meter signal-to-noise ratio s_m/n_m of unity seems a reasonable guess. However, in cases where the memory of the observer is particularly good or when, say, recording tapes are used, the value of s_m/n_m may be considerably less than 1. In other words, high sensitivity may also be secured by waiting for a sufficiently long period of time, even though the time-constant of the meter itself is not unduly long.

Use of a Gate

In order to build a very sensitive aural or metering scheme, one should not only have a very narrow filter or long time-constant, but in addition one should employ a gate. The effect of the gating device is to reduce the noise background by the "gating factor," i.e., the fraction of the time that the gate is open or transmitting. The gate, however, does not reduce the noise while the signal is being received. Hence the gating factor should not be used in computing the amount of modulation suppression of the signal in a linear rectifier due to the simultaneous presence of noise. When the gate has the same "on-off" period as the signal, it creates a periodic disturbance with the same period as the signal, and so might easily be confused with the latter in aural reception or in detection by a periodic metering scheme. However, there is the ingenious idea, due to Dr. W. H. Jordan (M.I.T. Radiation Laboratory), of opening the gate twice as often as the signal is repeated. The periodicity in the noise, arising from the gating, then has a fundamental equal to the second harmonic $2f_p$ of the signal period f_p, and so will not be heard at all when we listen to the fundamental of the PRF.

With ordinary video or cathode-ray presentation a gate is of no appreciable help as it does not affect the signal-to-noise ratio *while* the signal is being displayed. On the other hand, audio-detection examines a particular Fourier component, and compares the latter with the average noise background rather than just that in the presence of the signal. The insertion of a gate is here a very material help. The ability to obtain as great a sensitivity with the audio as with the video method is usually contingent upon the ability to use a rather narrow gate in the former, corresponding to a gating factor of the order 10^{-2} or so.

V. The Dependency of Visual and Aural Response on Filter Width and Pulse Duration Away from Match

We have mentioned in Section II that the result that optimum performance is obtained at match (i.e., when the filter response is the conjugate of the Fourier transform of the pulse) is independent of the spectral "shape" of the filter. However, when we wish to compare the performance at mismatch with that at match, the behavior depends somewhat on the shape of the pulse or the filter. In order to make explicit calculations possible, we henceforth assume throughout Section V (Part I), that the filter and pulse both have a gaussian structure* so that their amplitude-frequency characteristics are, except for constant factors, $\exp(-\omega^2/2\omega_b^2)$ and $\exp(-\omega^2/2\omega_g^2)$ respectively (where the origin of frequency ω is at the carrier). As the Fourier transform of a gaussian function is also gaussian, the pulse depends on time in a gaussian fashion, *viz.*, $\exp(-\omega_g^2 t^2/2)$. At match one has $\omega_b=\omega_g$, and we now investigate how much the performance is jeopardized by having mismatch, where $\omega_b \neq \omega_g$.

* *Note:* Our choice of the gaussian response is based on the observation that such a characteristic is a good approximation to actual ones in practise, except for frequencies well away from resonance. (See Sec. III, Part II.)

Varying Filter Width

Let us imagine first that the pulse width is given and the IF filter width is varied, so that ω_g is fixed and ω_b is changed. The power necessary to achieve a given minimum detectable signal is readily calculated and may be expressed simply in terms of the ratio of power necessary at mismatch to that required at match. The details of the analysis have already been carried through independently by Wang and Uhlenbeck at (M.I.T.) Radiation Laboratory, so we will omit them in this paper.

Figure 4 shows the necessary power as a function of $\lambda(\equiv f_b/f_g)$ for the case of aural reception. The curves for the linear detector are not the same as for the quadratic rectifier; namely, the shape of the curve in the linear case is a function of the signal-to-noise ratio $\sigma_0=(s_F/n_F)$ at match, which in turn depends on the minimum detectable signal. The dependence on σ_0 in the linear detector occurs because the value of σ_0 determines the amount of modulation suppression. If, however, the signal is very weak, the performance of the linear rectifier is the same as that of the quadratic except for a factor independent of λ. Comparison of the curves in Fig. 4 with the results for the visual case indicates that aural reception is somewhat less sensitive to mismatch.

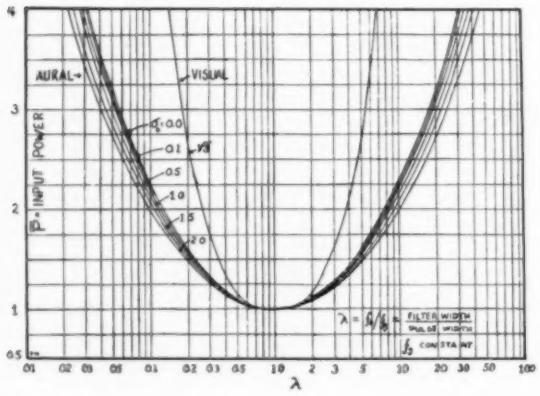


FIG. 4. Input power to the IF as a function of IF filter width at constant pulse width for aural reception of the minimum perceptible signal (Eq. 78b), with a linear rectifier. $\bar{P}$ is defined as the ratio of input power to input power at match. A gaussian pulse is assumed. The curve labelled $\sigma_0=0$ also applies for all values of σ_0 in the instance of a quadratic rectifier (Eq. 77b). A curve for visual reception is also included.

Varying Pulse Length

Let us now see what happens when the filter width is fixed and the pulse width is varied. In the case of visual reception the calculations show that the performance involves the filter width Δf_b and pulse length $\Delta T_g=1/\Delta f_g$ only through the product $\Delta f_b \Delta T_g$. Hence in the visual case it makes no difference whether we alter filter width or pulse length;* more power is always required when a condition of mismatch exists.

On the other hand, in aural reception the curve obtained by varying pulse length is of a quite different shape from that obtained by varying the filter width. The reason for this is that the quantities Δf_b and ΔT_g here do not enter purely in the product $\Delta f_b \Delta T_g$. Our proof of maximum efficiency at match applies only when Δf_b rather than ΔT_g is varied. If we define $\bar{\lambda}$ as the ratio $\Delta f_g/\Delta f_b$ of the frequency width of the pulse to its width at match, or, in other words, as the reciprocal of the ratio of the pulse length to its length at match, then the variation of necessary power with $\bar{\lambda}$ is shown in Fig. 5 for the linear and square detectors. The curves do not have minima at $\bar{\lambda}=1$. The necessary power is always diminished by lengthening the pulse. However, for extremely long pulses the necessary power approaches an asymptotic value which Fig. 5 shows is nearly achieved when $\bar{\lambda}$ is less than 0.1 or so. Here again the results for the linear detector depend on the signal-to-noise ratio σ_0, with the curve for $\sigma_0=0$ the same as for the quadratic rectifier. The upshot of Fig. 5 is that *long pulses are more easily heard, or detected by a meter than are short ones.*† Of course there are practical

* Within the approximations assumed in our criterion; see section 1 of part I.

† Possibly the reader is concerned with the physical reason why the dependence on pulse length is so different for the visual and aural cases. The explanation runs as follows: It is clear that in either case the power requirements will be increased if the pulse is made too short, for with very short pulses the Fourier spectrum becomes so wide that only a fraction of the energy is passed by the filter, with attendant wastage. In the visual case it is also wasteful to have the pulse too long, for with our "simple criterion," all that counts is the peak value of the pulse after filtering. Keeping the pulse on a very long time, i.e., much greater than the reciprocal of the frequency band width of the filter, will not appreciably augment this peak, but will increase the energy that has to be supplied, just because of the greater time duration. Hence with visual detection the pulse should be neither too long nor too short, and there is thus some particular pulse length at which the input power required to achieve a given signal-

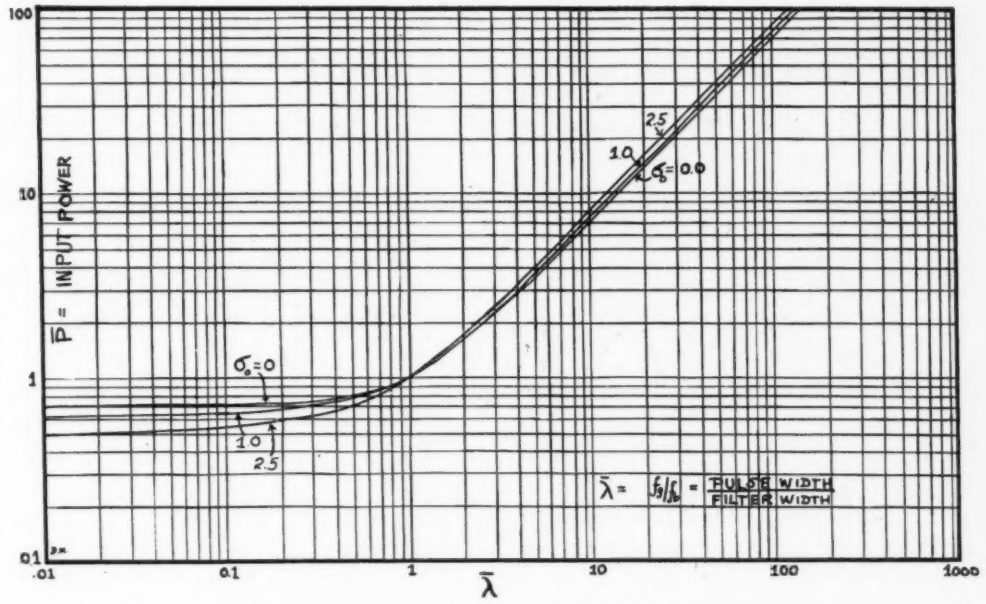


FIG. 5. Input power to the IF as a function of pulse width at constant filter width for aural reception of the minimum perceptible signal, Eq. (78a), with a linear rectifier and gated or ungated noise. The pulse has a gaussian shape. The width in frequency is here the reciprocal of the duration of the pulse. Again the curve $\sigma_0=0$ applies for all values of σ_0 in the case of a quadratic rectifier, Eq. (77a).

limits as to how much the pulse length can be increased, for if the pulse length is too great, there will be an overlapping of successive pulses, and the received intelligence would then be destroyed, not to mention the fact that mathematically our calculations are all predicated on the assumption of no overlapping.

VI. The Influence of PRF, Gating, and Audio-Filter Width on Sensitivity

Visual Reception

The dependence of the input power on PRF in visual reception can be dismissed rather summarily. Namely, with our "simple" criterion as to the minimum detectable signal, the only thing that counts is the energy per pulse. The mean power requirements are therefore directly proportional to the PRF, while the peak power will be independent of it. This is what one would expect, for with our "simple" criterion for measuring visual detection, which neglects the integrating effect of the eye (and brain), the only thing that is important is the instantaneous ratio of maximum transmitted signal energy to the noise background, and it is immaterial how many pulses are generated per second. Hence it is most economical to have the pulses spaced at very long intervals.

Neglect of the effect of integration in the visual case will have no bearing on the relative merits of the linear and square detector, and will not do much harm in considering questions of filter structure. On the other hand, in considering the

to-noise ratio is a minimum. In aural detection we are interested in the energy in one particular Fourier component. The total number of Fourier components of appreciable magnitude is of the order of the ratio of pulse period to pulse duration. Hence, if the pulse is lengthened the energy is distributed among fewer components, or, in other words, there is less wastage in the high harmonics, which are of no concern when we listen to the fundamental or a low harmonic. This effect counterbalances the fact that the greater duration of long pulses demands more power. Consequently, the power requirements approach asymptotically a limiting value when the pulse is very long, rather than increasing as in the visual case.

If, instead, the pulse length is kept constant and the filter is varied in width, it is clear that too wide a filter is wasteful, as it simply increases the total noise without improving the transmitted signal. At first thought it might seem that narrowing the filter a great deal would diminish signal and noise equally. However, the coherence of the various Fourier components of the signal (before rectification) as compared to the incoherence of the noise has the result that the rectified signal is weakened more than the noise by the narrowing, so that for an infinitely narrow filter the signal-to-noise ratio becomes zero, and consequently in Fig. 4 the power requirements become infinite as λ approaches zero.

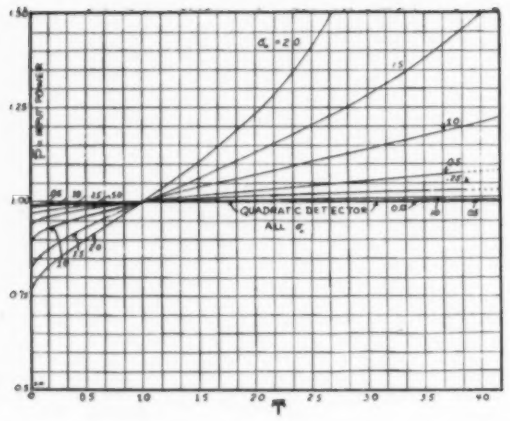


FIG. 6. Input power to the IF as a function of pulse period ($\bar{T}=T/T_0$) for aural reception at match, of the minimum perceptible signal, with *ungated* noise and linear rectifier, Eqs. (77a, 77b). Here $\bar{T}$ is the ratio of the pulse period to the pulse period at match.

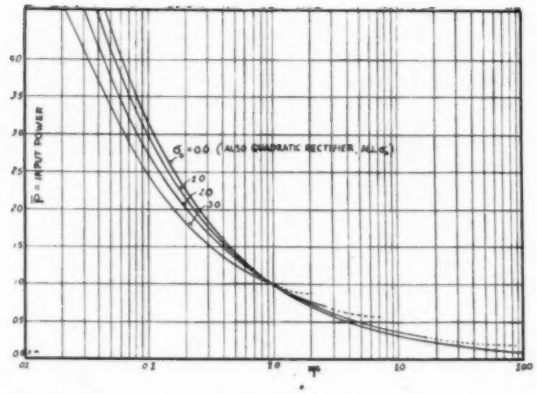


FIG. 7. Input power to the IF as a function of the pulse period $\bar{T}=T/T_0$ for aural reception of the minimum perceivable signal, Eqs. (77a, b), at match with *gated* noise and a linear rectifier. The curve for $\sigma_0=0$ also applies for all σ_0 in the case of a quadratic rectifier, Eq. (80).

dependence of visual efficiency on PRF it is certainly important to include the integration effects which we neglect. The theory of North and of Uhlenbeck indicates that at least over a certain range of PRF the visibility of a pulse is proportional *to the square-root* of the PRF. This has been borne out by the extensive experiments of Lawson (M.I.T. Radiation Laboratory). Hence the effect of integration is to make the average power necessary to achieve a minimum detectable signal proportional to the square root rather than to the first power of the PRF, and to make the peak power inversely proportional to the square root of the PRF.

Audio-Detection—No Gate

In the audio case the precise formulation of the minimum detectable signal presents itself in a more clear-cut fashion inasmuch as our criterion in terms of a particular Fourier component already includes the effect of integration. There still is, to be sure, a question as to what is a proper numerical value of the critical threshold.

Because the gating factor $G=\tau_G/T$ is directly proportional to the PRF, we must be careful to distinguish between those cases for which gating is or is not present in aural presentation.

First, let us assume that there is no gate. With no gate, theory shows (see Eqs. (67), (68)) that for the square-law device, the minimum detectable signal is independent of the PRF and at match is a function only of the input power to the IF and of the IF filter width. This behavior can be explained in the following way. For a given energy per pulse and a given IF width, the longer the interval between pulses the less power consumed, but the power is less effective because the Fourier components are more closely spaced, and hence there is less energy in any one component. More precise analysis shows that with a quadratic rectifier these two effects just cancel, so that for a *given pulse duration* and IF filter width, the input power to the IF required to achieve a particular value of the minimum detectable signal-to-noise ratio s_a/n_a is not affected by the PRF.* This conclusion holds quite irrespective of whether the filter is matched or the pulse is gaussian in shape. For the linear rectifier this independence of PRF holds only in the limiting case of a very small value of s_F/n_F. Actually, for non-vanishing s_F/n_F the performance with a linear detector is a slowly varying function of the pulse repetition frequency, the input signal being more easily detectable when the PRF is increased as is shown in Fig. 6. The curves of Fig. 6 are drawn on the assumption of a gaussian pulse with matched filter, but the trend with PRF will also be the same in other cases, as the exact filter structure merely determines the precise degree of modulation suppression, and so is reflected mainly in the absolute power scale.

* We are speaking here and elsewhere of mean power. The peak power is correspondingly inversely proportional to PRF.

Insertion of a Gate

The conclusions of the preceding paragraph also apply as long as the PRF is varied but the gating factor (percentage of time the gate is open) is held constant, as then the gate merely changes the absolute sensitivity of the instrument. Physically, however, one usually varies the PRF, keeping constant the absolute time the gate is open, rather than percentage time. Then the dependence on the PRF is more complicated and needs discussion.

When a gate is used and its duration is held constant, it turns out that less input power is required if the PRF is decreased. This is because the gating factor is decreased, which in turn means that the background noise is lowered with respect to the input signal level, and thus we are able to obtain greater sensitivity (see Eq. (79)). Theory shows (cf. Eq. (80)), that with a quadratic rectifier and a gate of constant duration, the mean power necessary to achieve a given signal-to-noise ratio is proportional to the square root of the PRF. (The peak power is correspondingly inversely proportional to the square root of the PRF.) This result is general and does not require a particular pulse or filter shape, such as, for instance, the gaussian. Curves giving the input power as a function of PRF are shown in Fig. 7.* Here it is assumed that the filter is matched and gaussian, but in other cases, the curves will not be appreciably different, in fact, not at all in the case of the quadratic rectifier.

Instead of keeping the gate fixed and varying the PRF, one may also be interested in what happens when the PRF is constant and the width τ_G of the gate is altered. The variation of the necessary input power with τ_G under these conditions is shown in Fig. 8. Obviously, narrowing the gate has the same effect as narrowing the earphone or meter filter. The abscissa of Fig. 8 can thus be equally well taken as the gate or the earphone width, depending on which one is interested in varying. For the quadratic detector, where questions of modulation suppression do not enter, the power is proportional to the square root of the gate width (or of the audio-filter width). This behavior for the quadratic rectifier, which also applies to the linear detector with very weak signals, does not require any specialization, but the curves of Fig. 8 for the response of linear rectifiers with non-vanishing values of σ_{00} apply accurately only to a matched, gaussian pulse. The plots for other filters, however, will not be appreciably different.

When we seek to build up the efficiency by narrowing the gate, and when the latter precedes the IF filter, there are limits beyond which we cannot go. For if we were to decrease the gating factor to too narrow a relative time interval, it would be impossible to detect a target which was moving rapidly enough to pass through the equivalent space range of the gate between two successive scans, since the echoes returning from the target might then arrive at the receiver during a period in which the gate were shut. Apart from these considerations, gating under such circumstances would result in a distortion of the pulse by shearing it fore and aft (an effect not considered in our calculations), so that the condition for best match for the filter would be changed. If a gate of short duration is used at all, it is better to place it after the IF filter and corresponding rectification. As noted by North, it is possible

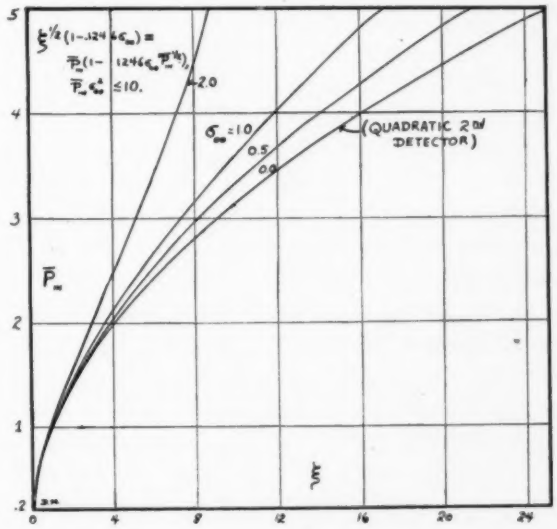


FIG. 8. Input power to the IF for aural or aperiodic meter reception of the minimum detectable signal at match using a linear rectifier, as a function of ξ, where ξ represents (1): duration of the gate, or (2) width of the audio-filter, or (3) frequency width of pulse (or IF filter), where gaussian pulses and filters are assumed. The subscript 00 refers to matching the IF to the pulse.

* The curves in Fig. 7 for the linear detector do not extend beyond a certain value of $\bar{T}$ for any particular σ_0 ($=s_F/n_F$ at match). The reason for this is mathematical and not physical, for the formulas involved are not reasonably correct beyond the point indicated.

under these circumstances to use a gate of infinitesimal length. The item of adequate space coverage of targets is then secured by the finite length of the pulse. Of course, the ideal system would be to have an infinitesimal gate come right at the maximum of the rectified pulse, but uncertainty in range will in general make this impossible. The great advantage of visual detection is that the eye is, so to speak, its own properly timed gate. In attempting to discern a signal, it compares a pip or signal maximum only with the noise background in its immediate vicinity in time. North's rather elegant concept of an infinitesimal gate has the advantage that if the gate is inserted at just the right time, i.e., at the maximum of the pulse, the criterion for a minimum detectable signal is obviously similar for both aural and visual detection, for then in either case we are concerned only with the ratio of signal-to-noise energy at the very peak of the signal. In practise, however, an ideally-timed, infinitesimal gate seems too much of a simplification, for reasons already stated, but still it is a useful concept,* delineating a limiting behavior.

VII. Relative and Absolute Sensitivities of Visual and Aural Detection: Behavior at Match

As already mentioned many times, we say that the filter is "matched" if it is the conjugate of the Fourier transform of the pulse. The study of the matched case is particularly fundamental for three reasons (a), it is always a measure of best achievable performance, (b), it furnishes a convenient norm with which to compare results in other cases, and (c) it is particularly simple from the mathematical standpoint, especially in the visual case.

Hence in our attempts to calculate absolute sensitivities numerically in the end of this present section, we shall always assume that a condition of match exists. When in the present section we talk about the dependence on pulse length at match, the reader should be careful not to confuse this type of dependence on pulse length with that treated in the preceding section, where the filter width was assumed constant. We now consider the type of variation for which, when the pulse length is altered, the filter is correspondingly changed so as to be always at match.

With *visual perception* (on the deflection-modulated oscilloscope), the theory of Section III, Part II shows that the results obtainable at match are completely independent of the pulse shape (on the assumption of the "simple" criterion, of course). For the square detector one has, at match, from (67) and (73), the very simple result

$$(s_v/n_v)_{\text{quad}} = P_0T_0/W_0,$$

where W_0 is the power density of input noise, per unit frequency range, in the spectral region of the IF filter, P_0 is the input power, and T_0 is the pulse repetition period. For the linear detector the corresponding formula is, from (67) and (75)

$$(s_v/n_v)_{\text{lin}} = 0.957\frac{P_0T_0}{W_0} \times\left(1-0.1526\left(\frac{P_0T_0}{W_0}\right)^{\frac{1}{2}}\right), \quad P_0T_0/W_0 \leq 10.$$

Since P_0 and T_0 occur only in the product P_0T_0/W_0, the only thing that counts is the input energy per pulse, *viz.*, $E=P_0T_0$, Eq. (63). This is because our "simple" criterion for visual performance neglects the integrating effect of the eye, as already discussed in detail in Section VI, Part I.

Our conclusion that in visual detection the performance at match is independent of the pulse length is doubtless too much of an idealization, as the "simple" criterion takes no cognizance that the visibility of a pip is probably somewhat conditioned by its width. However, widening or narrowing of the pulse and pip will also be accompanied at match by fattening or shrinking of the "blades of grass" due to noise. Because of this fact, the invariance of pulse structure pre-

* In our mathematical section (Part II) we do not treat the infinitesimal gate. However, we can apply in a rather straightforward manner a fluctuation analysis, applied to pulses which are the delta-functions passed by this infinitesimal gate. If we assume that the gate is properly timed, and that a second gate is present to cancel out the discrete noise in the fundamental of the PRF, then it can be shown that if $(s/n)_F$ is given, $(s/n)_{a\ \text{infin gate}}$ exceeds the visual ratio $(s/n)_v$ by a factor $(T\Delta f_e)^{-\frac{1}{2}}$, where $(T\Delta f_e)^{-1}$ is the ratio of PRF to audio- (or meter) filter width. Even with perfect timing the performance is not appreciably improved by making the gate much shorter than the pulse. In other words, our later Eq. (c) of Section VII, with $\tau_0 \sim 1/\Delta f_b$, gives a value of $(s/n)_a$ of the same order as that which we have just estimated for an infinitesimal gate.

TABLE I.

Output Ratios			Input Ratios σ_0			
			Quadratic 2nd detector		Linear 2nd detector	
	Physically minimum Detectable signal		$\Delta f_a = \Delta f_e$ 50 c.p.s.	Δf_e 5 c.p.s.	Δf_e 50 c.p.s.	Δf_e 5 c.p.s
(Visual)	0.0 db	(pessimistic)	$\sigma_0 =$	0.0 db		+1.0 db
(s_v/n_v)	− 5.0			−2.5		−1.7
	−10.0	(optimistic)		−5.0		−4.4
(aural)	+10.0	(pess.)	+13.5	+ 8.5	>13.5	+11.2
(s_a/n_a)	0.0	(opt.)	+ 8.5	+ 3.5	>13.5	+ 4.8
no gate						
(aural)	+10.0	(pess.)	+ 2.0	− 3.0	+ 2.9	− 2.5
(s_a/n_{ag})	0.0	(opt.)	− 3.0	− 8.0	− 2.5	− 7.8
gate						
	Δf_m	Time constant $=2/\Delta f_m$				
periodic metering	2 c.p.s.	1 sec.	$\sigma_0 =$ − 6.7db	− 9.3	− 6.3	− 9.0
$(s_m/n_m) = 0.0$ db	1/5	10	− 9.3	−12.0	− 9.0	−11.5
gated noise, lin. rect.	1/10	20	−10.0	−12.4	− 9.8	−12.2
aperiodic metering	2	1	−10.0		− 9.7	
$(s_m/n_m) = 0.0$ db	1/5	10	−15.5		−15.2	
	1/10	20	−16.5		−16.3	

The equivalent rectangular pulse is 1.41 μsec. long. $T_0 = 10^{-3}$ sec., $\tau_G = 5 \times 10^{-4}$ sec.; $f_b = 4 \times 10^5$ c.p.s.

dicted by our "simple" criterion is probably not too much of a sin, unlike the conclusions which it yielded concerning the dependence of visibility on PRF.

With *aural reception* the performance at match is no longer independent of pulse shape or length. Instead, the power required to reach a minimum detectable signal-to-noise ratio is less, the longer the pulse. The dependence of this power on pulse duration turns out to be similar to that on gate length studied in the previous section. As long as the shape of the pulse remains constant, and only its time-scale is changed, the power demanded is inversely proportional to the square root of the length of the pulse, if the detector is quadratic or if it is linear and the signal is weak. The variation of the necessary power, with pulse length in arbitrary units, is shown in Fig. 8 for the quadratic detector and for the linear rectifier with various thresholds.

The reason for this difference in behavior between the visual and aural cases as regards the dependence on pulse length is that the expression to be maximized under optimum filter design is different in the two cases, and it can be shown that in consequence the maximum in one case has the value unity quite irrespective of the shape of the pulse, whereas in the other the maximum depends on the assumptions concerning the structure and length of the pulse (cf. Eqs. (52), (55), and remarks in second paragraph after Eq. (56)). Physically the reason that increasing the pulse length is helpful in aural but not visual detection is that it lowers the fraction of energy wasted in irrelevant higher harmonics of the PRF in aural reception, whereas in the visual case it prolongs the maximum without raising it, which with our "simple" criterion is no help.

Numerical Calculations of Sensitivities

Knowing the values of the physically minimum detectable signal-to-noise ratios for the various types of presentation, oscilloscope, ungated, or gated aural reception, and use of a meter, we can calculate and compare the relative and absolute sensitivities of these methods. To do so we use the results obtained in Part II, and quote some of the final formulas here. For visual reception we have the relations

$$(s_v/n_v)_{\text{quad}} = \sigma_0^2; \qquad (s_v/n_v)_{\text{lin}} = 0.957\sigma_0^2(1 - 0.1526\sigma_0), \quad \sigma_0 \leq 10^{\frac{1}{2}}, \tag{a}$$

which follow from (73) and (75), at match. For aural reception, the necessary formulas are, from (48a, b) and (49a, b) and Section IV (b), Part II

for *ungated* noise,

$$(s_a/n_a)_{\text{quad}}=\frac{0.633\sigma_0^2}{T_0(\Delta f_e f_b)^{\frac{1}{2}}};$$

$$\left(\frac{s_a}{n_a}\right)_{\text{lin}}=\frac{0.616\sigma_0^2}{T_0(\Delta f_e f_b)^{\frac{1}{2}}}(1-0.1246\sigma_0), \qquad \sigma_0\leq 10^{\frac{1}{2}}. \tag{b}$$

Here Δf_e is the width of the critical ear or audio-filter, whichever is narrower. When the background is gated, (b) becomes

$$(s_a/n_{aG})_{\text{quad}}=\frac{0.633\sigma_0^2}{(T_0\tau_G\Delta f_e f_b)^{\frac{1}{2}}};$$

$$(s_a/n_{aG})_{\text{lin}}=\frac{0.616\sigma_0^2}{(T_0\tau_G\Delta f_e f_b)^{\frac{1}{2}}}(1-0.1246\sigma_0), \qquad \sigma_0\leq 10^{\frac{1}{2}}. \tag{c}$$

The expression for the periodic meter is obtained from (97) and is

$$(s_m/n_m)=1.156\frac{[1+\frac{1}{2}(\Delta f_m/\Delta f_a)^2]^{\frac{1}{2}}}{(\Delta f_m/\Delta f_a)^{\frac{1}{2}}}\times\{(s_a/n_{a(G)})^2-0.1526(s_a/n_{a(G)})^3\}, \qquad s_a/n_{a(G)}\leq 10^{\frac{1}{2}}, \tag{d}$$

where the second detector that is used may be either linear or quadratic. The parenthesis around the subscript G in (d) means that the appropriate value of s_a/n_a is (b) or (c) according as the noise is ungated or gated.

The quantity Δf_m appearing in (d) is twice the equivalent rectangular filter width of the meter. (The reason for the factor two is that our general theory contemplates a filter extending symmetrically on the two sides of a resonance center, whereas the low pass filter involved in a meter is unilateral.)

To illustrate and compare the absolute sensitivities for the various methods Table I has been prepared for the following conditions of operation. First, we have assumed a gaussian pulse, to which the filter is matched. The width of the equivalent rectangular pulse is $7.1\cdot 10^5$ c.p.s., $f_b=4\cdot 10^5$ c.p.s. from (60), and the width between half-power points is $\frac{3}{4}$ Mc. Furthermore, we have taken $\tau_G=5\cdot 10^{-6}$ second (the total duration of a *double* gate; see Sec. IV and VI), and $T_0=10^{-3}$ sec., corresponding to a PRF of 1000 c.p.s. Two different audio-filters are used: $\Delta f_e=50$, $\Delta f_e=5$ cycles. Three distinct meter filters are also employed: $\Delta f_m=2$, $\frac{1}{5}$, $\frac{1}{10}$ cycles (actual width 1 cycle, $\frac{1}{10}$, 1/20 cycles), each with the same audio-filter $\Delta f_e=50$ or $\Delta f_e=5$ c.p.s.

Here σ_0 equals (s_F/n_F) at match, and is given in db in the appropriate boxes of the table. We have assumed that $s_m/n_m=1$, but this is only a guess at present, for want of more complete experimental data. From this table we see that the linear detector is always a bit worse than the quadratic, though the difference for weak signals is small. We notice also that under the above conditions of operation, aural reception is about as good as visual, and in some cases surpasses it, although the psychological strain of aural work may add as much as 5 db to these results. The meter, of course, compares favorably on the assumption that $s_m/n_m=1$ as a minimum detectable signal. Narrowing the audio-filter by a factor 10 increases the sensitivity by 5 db. Eaton and Wolff (RCA Princeton, in an unpublished paper (Report PTR-7C) completed June 4, 1943), using the aperiodic method, with time-constants of the order of 10 seconds, obtained in the Laboratory a 20-db improvement over visual detection by the (deflection-modulated) oscilloscope if $s_v/n_v=1$. This is about the performance that our theory indicates. Namely, if we take $\Delta f_m=\frac{1}{5}$ c.p.s., corresponding to a time-constant of 10 seconds, and take a single gate of about 1.5 μ-seconds duration, as they used a gate comparable with the pulse length (rather than the longer and double gates assumed in our tables), our formulas yield an 18.5 db gain in sensitivity over the "pessimistic" oscilloscope reading. We cannot, however, emphasize too much that the efficacy of the various metering methods is contingent upon the feasibility of using a long time-constant, and this will depend on tactical considerations (of Sec. IV).

PART II. MATHEMATICAL SECTION

I. Introduction: Signal and Noise in the IF Filter

The pulse-modulated carrier enters the RF stages and the carrier frequency is shifted to the lower frequency, (IF), by means of the first detector or mixer, as it is often called. We assume

that this displacement of the carrier, as well as the performance of the RF stages, is essentially a linear process, resulting in no appreciable distortion of the incoming wave. Our real problem, then, is with the IF and subsequent stages of the receiver. The noise entering the IF arises primarily in the mixer, or in the case of artificial noise, comes, of course, through the RF. The noise and signal are filtered before entering the second detector. After rectification, the mixture is filtered by the video- or an audio-filter, amplified, and passed finally to an oscilloscope, earphones, or to a meter.

The signal entering the IF stages has the following forms:

$$f_I(t) = \sum_j \sqrt{2} s_I(t - t_j) \cos \omega_0 (t - t_j), \qquad (1)$$

where $\omega_0 = 2\pi f_0$ and f_0 is the intermediate frequency. Equation (1) represents the incoming wave as a succession of pulses, the envelope of any one of which is described by $\sqrt{2} s_I(t - t_j)$. The different values of t_j differ by the pulse repetition period T. A single disturbance is expressed by $\sqrt{2} s_I(t) \cos \omega_0 t$, with $t_j = 0$ for convenience. We shall also assume that the overlapping between consecutive pulses is negligible, so that in determining (1) it is necessary to consider only one value of t_j at a time. For all purposes this is a very good approximation. The Fourier transform of a single pulse is then

$$S_I'(\omega') = \int_{-\infty}^{\infty} \sqrt{2} s_I(t) \exp(-i\omega' t) \cos \omega_0 t \, dt, \qquad \omega' = 2\pi f'. \qquad (2)$$

Note that from (2) it follows automatically that $S_I'(\omega') = S_I'(-\omega')$,* and $S_I'(\omega')^* = S_I'(-\omega')$, where the star (*) denotes the complex conjugate. The corresponding relation to (2) is

$$\sqrt{2} s_I(t) \cos \omega_0 t = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_I'(\omega') \exp(i\omega' t) d\omega'. \qquad (3)$$

As well as amplifying, the IF stages act like a filter, whose *amplitude* frequency response we denote by $G_{IF}'(\omega')$. For the time being we will not restrict the generality of our treatment by specifying $G_{IF}'(\omega')$ explicitly. The filtered signal that leaves the IF is then

$$s_F'(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_I'(\omega') G_{IF}'(\omega') \exp(i\omega' t) d\omega', \qquad (4)$$

where we must have $G_{IF}'(-\omega') = G_{IF}'(\omega')^*$ in order that we have a real output for a real input. The primes are dropped from the notation for the various filter characteristics when the origin for frequency is taken at the carrier f_0 rather than at the absolute zero of frequency: thus $\omega = \omega' - \omega_0$, $S_I'(\omega' - \omega_0) = S_I(\omega)$, $G_{IF}'(\omega' - \omega_0) = G_{IF}(\omega)$. Then with the substitution $\omega = \omega' - \omega_0$ we may write (2) without appreciable error as

$$S_I(\omega) = 2^{-\frac{1}{2}} \int_{-\infty}^{\infty} s_I(t) e^{-i\omega t} dt, \qquad (5)$$

since we can disregard the part of the integrand with the oscillatory factor $\exp(-2i\omega_0 t)$, whose effect is negligible because of the rapid variation. Equation (4) can be written

$$s_F'(t) = 2 \times \text{Real Part of}$$

$$\left\{ \frac{\exp(i\omega_0 t)}{2\pi} \int_{-\omega_0}^{\infty} S_I(\omega) G_{IF}(\omega) e^{i\omega t} d\omega \right\}. \qquad (6)$$

We can express the filtered wave as

$$s_F'(t) = \sqrt{2} s_F(t) \cos(\omega_0 t - \epsilon), \qquad (7a)$$

where

$$\sqrt{2} s_F(t) = 2 \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} S_I(\omega) G_{IF}(\omega) e^{i\omega t} d\omega \right|. \qquad (7b)$$

The phase factor ϵ is really a function of t but is so slowly varying that to all intents and purposes it can be regarded as constant. It is to be noted that the integrand of (7b) has a single resonance center at $\omega = 0$, whereas that of (4) has resonances at $\omega = \pm\omega_0$. In extending the lower limit from $-\omega_0$ to $-\infty$ in (7b) it is to be understood that $G_{IF}(\omega)$ is to be taken as very small for $\omega < -\omega_0$, whereas the literal definition $G_{IF}(\omega) = G_{IF}'(\omega' - \omega_0)$ would imply that $G_{IF}(\omega)$ had a resonance at $-2\omega_0$ also.

Equation (7b) shows that the amplitude $s_F(t)$ after filtering can be computed, as we would expect, just as though we ignored the carrier entirely, and applied the Fourier analysis directly to the modulation factor $s_I(t)$ of the incoming wave. We call $s_I(t)$ a modulation factor rather than an envelope because in evaluating (7b) it is essential to take account the algebraic sign as well as the modulus of $s_I(t)$.

We will assume that the background noise entering the IF is sensibly uniform over the filter width, having a constant spectral density W_0 (power/frequency). This is really no assumption at all, as the spectral variations in the noise over the IF filter width are usually of little account. The mean square noise power transmitted by the IF stage is then

$$n_F^2 = \frac{1}{2}\frac{W_0}{2\pi}\int_{-\infty}^{\infty} |G_{IF}'(\omega')|^2 d\omega' = \frac{W_0}{2\pi}\int_{-\infty}^{\infty} |G_{IF}(\omega)|^2 d\omega, \quad (8)$$

where n_F is the r.m.s. noise voltage entering the rectifier.

II. Quadratic and Linear Rectification of Signal and Noise

We introduce now the correlation function $R(t)$ for the noise and signal output of a linear or nonlinear device, such as a filter or rectifier, in terms of the power spectrum of the output $W(\omega)$. The correlation $R(t)$ is defined as the mean value of the product of the noise and signal amplitudes at times separated by an interval t, averaged over all initial phases. It can be shown[6–10] also that the correlation function is the Fourier transform of the output power spectrum, and *vice versa*. We have thus the following general relations:

$$W(\omega) = 4\int_0^{\infty} R(t)\cos\omega t\, dt, \quad (9)$$

and

$$R(t) = \frac{1}{2\pi}\int_0^{\infty} W(\omega)\cos\omega t\, d\omega. \quad (10)$$

For noise leaving the IF filter we have the correlation function $\psi(t)$, *viz.*

$$\psi(t) = \frac{W_0}{4\pi}\int_{-\infty}^{\infty} |G_{IF}'(\omega')|^2 \cos\omega' t\, d\omega'. \quad (11)$$

A comparison of (11) and (8) shows immediately that $n_F^2 = \psi(0)$, so that $\psi(0)$, or in abbreviated form ψ, is the mean square noise voltage passed by the IF filter, which is centered about f_0. Then with the help of (11) we can write with negligible error

$$\psi(t) \equiv \psi_0(t)\cos\omega_0 t \equiv \psi(0)r(t) \equiv \psi(0)r_0(t)\cos\omega_0 t. \quad (12)$$

The relations (12) define the normalized correlations $r(t)$ and $r_0(t)$, where

$$\psi_0(t) = \frac{W_0}{2\pi}\int_{-\infty}^{\infty} |G_{IF}(\omega)|^2 \cos\omega t\, d\omega, \quad (13)$$

since the contribution to the integral when $\omega < -\omega_0$ is ignorable. Note that $r_0(0) = r(0) = 1$.

(a). *Noise and Signal Output of a Quadratic Rectifier*

Let us now pass the signal from the IF with its accompanying noise through a square-law detector. It can be shown[10] that the correlation function $R_N(t)$ before averaging over the phases of the modulation $s_F(t)$ is

$$R_N[t, t_0, t_0'] = \eta^2\{\psi^2 + b_1^2 b_2^2\psi^2 + 2\psi(t)^2 + 4b_1 b_2\psi\psi(t) + b_1^2\psi^2 + b_2^2\psi^2\}, \quad (14)$$

where

$$b_1 = \sqrt{2}s_F(t_0')\cos\omega_0 t_0/\psi^{\frac{1}{2}}; \quad b_2 = \sqrt{2}s_F(t_0'+t)\cos\omega_0(t_0+t)/\psi^{\frac{1}{2}}, \quad (15)$$

and η is a constant depending on the dynamic characteristic of the particular rectifier. Assuming that the envelope $s_F(t_0')$, etc., is a slowly varying function of the time, we may treat it as essentially constant, and hence uncorrelated with $\cos\omega_0 t_0$, etc., when averaging over the phases of the carrier. Then analytically t_0 and t_0' are independent. Now in the final form of the correlation function the terms in ψ^2, $b_1^2\psi^2$, $b_2^2\psi^2$ contribute only to the d.c. Further, the term in $\psi(t)^2$ with the help of (12) may be resolved into $\psi(t)^2 = \frac{1}{2}\psi_0(t)^2(1+\cos 2\omega_0 t)$, the latter por-

[6] N. Wiener, Acta. Math. **55**, 117 (1930), and later, independently, A. Khintchine, Math. Ann. **19**, 604 (1934).
[7] G. I. Taylor, Proc. Lond. Math. Soc. Sec. 2, **20**, 196 (1920), and Proc. Roy. Soc. **164**, 476 (1938).
[8] S. O. Rice, Bell Sys. Tech. J. **23**, 282 (1944), and **24**, 46 (1945). These two papers contain an excellent and quite comprehensive treatment of a large number of noise problems.
[9] J. H. Van Vleck, "The Spectrum of Clipped Noise" (Harvard Radio Research Laboratory, Report 411–51, July 21, 1943). Available at Harvard War Archives, Littauer Building, Harvard University.
[10] D. Middleton, "Note on The Theory of Square Law Rectification of Modulated Carrier in Presence of Noise." (Harvard Radio Research Laboratory, completed July 12, 1944.) Available at Harvard War Archives, Littauer Building, Harvard University.

tion of which contributes to the spectral region centered about $2f_0$, well removed from the region of our interest. Accordingly, the low frequency correlation function derived from (14) becomes in final form[11]

$$R_{LF}(t)=\eta^2\{\psi_0(t)^2+2\psi_0(t)\overline{\langle s_F(t_0')s_F(t_0'+t)\rangle}+\overline{\langle s_F(t_0')^2 s_F(t_0'+t)^2\rangle}\}, \quad (16)$$

where the bars indicate the average over t_0'. The first term of (16) represents noise alone, the second is the contribution to the continuum from the cross-modulation of the signal and the noise, and the third is the signal output alone. In both visual and aural reception we can in general disregard the middle, or second, term. The reasons behind this, however, are quite different in the two cases. For aural performance the signal is on such a small fraction of the pulse period that the contribution of the cross-term to the continuum is negligible in comparison with the noise and the signal terms. In the video case it is still true that these cross-terms are present only a small fraction of the time, but now it is that fraction of the time that interests us. However, the response is always a monotonically increasing function of s_F/n_F, which is physically obvious, since increasing the signal must always lead to an improvement in reception. The best performance is thus obtained by maximizing s_F/n_F. We may then take as our definition of signal to noise ratio the r.m.s. signal (envelope) at that point in the receiver circuit where we perceive it, to the r.m.s. value of the corresponding noise background, omitting the cross-modulation terms.

In the absence of signal the low frequency correlation function is, from (16),

$$R_{LF}(t)=\eta^2\psi(0)^2 r_0(t)^2=\eta^2\psi_0(t)^2, \quad (17)$$

and from (10) it follows when $t=0$ that the low frequency mean power output is

$$R_{LF}(0)=\eta^2\psi(0)^2\equiv\eta^2 n_v^2=\eta^2 n_F^4, \quad (18)$$

where n_v is the mean square noise voltage in the video, provided the effective band width of the IF is less than that of the video stages, so that the low frequency output after detection is not appreciably affected by the video response. The spectrum of the low frequency noise from the quadratic rectifier may be obtained from (9) with the help of (17), and is*

$$W_{LF}(\omega)=4\int_0^\infty R_{LF}(t)\cos\omega t dt. \quad (19)$$

We have seen that (18) gives an expression for the video noise background. In aural reception we are interested in detecting a single component against a relatively narrow section of the spectrum, so that it is the spectral level of the noise near $f=0$ that concerns us. From (19) this is

$$W_{LF}(0)=4\int_0^\infty R_{LF}(t)dt. \quad (20)$$

The mean square *aural* background noise n_a^2 is rigorously

$$n_a^2=\frac{1}{2\pi}\int_0^\infty W_{LF}(\omega)\,|G_a(\omega)|^2 d\omega, \quad (21a)$$

where $G_a(\omega)$ is the amplitude response of the audio-filter, but we can use the approximation

$$n_a^2=\frac{W_{LF}(0)}{2\pi}\int_0^\infty |G_a(\omega)|^2 d\omega. \quad (21b)$$

It is convenient to express the band width of physical filters, whether measured from the half-power or other points of the response, in terms of a rectangular filter whose spectral ordinate is equal to the maximum response of the particular filter in question, and whose width is such that both devices transmit equal noise power, i.e., such that the areas under the respective responses are the same. (We assume filter characteristics with but a single maximum; the criterion may be generalized for more involved spectral shapes.)

[11] Equations (14) and (16) are derived on the assumption of small signal quadratic detection, arising from the square-law character of the dynamic response in the neighborhood of the operating point. It often happens that "large-signal" rectification occurs, where now the rectifier may have essentially a half-wave quadratic response. However, the results for the low frequency correlation are proportional, Eq. (16) being 4 times as great as the corresponding expression for half-wave detection. This is easy to see when we observe that only the envelope of the incoming, narrow-band disturbance is reproduced, albeit squared, in the low frequency output, exclusive of d.c. The factor of proportionality follows since in the latter case but half the input wave is transmitted.

* *Note:* After rectification there exists no carrier frequency in the spectral region with which we are concerned; here ω refers to absolute frequency, and not to frequency measured with respect to ω_0.

If $G_k(f)$ is the response of the physical filter, then the effective band width Δf_k of the equivalent rectangular apparatus becomes

$$\Delta f_k = \int_{-\infty}^{\infty} |G_k(f)|^2 df \Big/ |G_k(f)|_{\max}^2. \quad (22)$$

In all our definitions we will assume that $|G_k(f)|$ is normalized so that $|G_k(f)|_{\max}=1$. Then (21b) may be written still more simply as

$$n_a^2 = W_{LF}(0)\Delta\omega_e/2\pi = W_{LF}(0)\Delta f_e, \quad (23)$$

where in this instance Δf_e is the width of the earphone or audio-filter, if the band is very narrow, or the actual band width of the ear, if the audio-filter is wider. (See Fig. 1 for curves of (Δf_e) critical.)

(b). *Noise Output of a Linear Rectifier*

For the linear detector the low frequency part of the correlation function exclusive of d.c. for noise alone after rectification may be shown to be[12]

$$\begin{aligned} R_{LF}(t) &= \frac{\beta^2\psi(0)}{2\pi}\{{}_2F_1(-\tfrac{1}{2},-\tfrac{1}{2};1;r_0(t)^2)-1\}, \\ &= \frac{\beta^2\psi(0)}{8\pi}\{r_0^2+r_0^4/16+r_0^6/64 \\ &\qquad +25r_0^8/4096+\cdots\} \quad (24) \\ &= \frac{\beta^2\psi(0)}{8\pi}\sum_{n=0}^{\infty}\left\{\frac{(2n)!}{2^{2n}n!(n+1)!}\right\}^2 r_0^{2n+2}, \\ &= \frac{\beta^2\psi(0)}{2\pi^2}\{4E(r_0)-2(1-r_0^2)K(r_0)-\pi\}, \end{aligned}$$

where ${}_2F_1$ is the (Gaussian) hypergeometric function, and E and K are complete elliptic integrals of modulus r_0, with $r_0(t)$ specified by (12). The mean square video noise voltage n_v^2 without signal follows at once from (24), with the aid of

$${}_2F_1(\alpha,\beta;\gamma;1)=\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)/\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta), \quad R(\gamma)>0, \quad R(\gamma-\alpha-\beta)>0.$$

We have

$$n_v^2 = \beta^2\psi(0)\left[\frac{2}{\pi^2}-\frac{1}{2\pi}\right]=0.0435\beta^2\psi(0). \quad (25)$$

The low frequency continuum to be used in (21), or in (23), is immediately obtained from (19) and (24), to wit:

$$W_{LF}(\omega)=\frac{\beta^2\psi}{2\pi}\sum_{n=0}^{\infty}\left\{\frac{(2n)!}{2^{2n}n!(n+1)!}\right\}^2 \times\int_0^{\infty} r_0(t)^{2n+2}\cos\omega t dt. \quad (26)$$

(c). *Signal Output of a Quadratic Rectifier in Visual and Aural Reception*

As can be seen from (16), the low frequency signal output of a quadratic detector is independent of the background noise, if we neglect the cross-modulation term; (see reference 11). In visual reception it is the mean value of the area under one cycle of the IF carrier, or equivalently, the envelope of the pulse, that is important. The signal displayed appears as a single pulse $s_v(t_0'-t_j)$. We have assumed that the modulation is very slow compared to f_0, and so $s_v(t_0'-t_j)$ is essentially constant in time, t, when it appears in $R_N(t, t_0, t_0')$, as in integrating over t_0 we are dealing with times of the order of $1/f_0$. In order to emphasize the approximate constancy of $s_v(t_0'-t_j)$ in this integration we denote the argument by $t_0'-t_j$ rather than by t_0-t_j. However, it is not legitimate to ignore the time dependent nature of $s_F(t)$ and $s_v(t)$ when later we consider signal variations over times of the order of ω_p^{-1}, where ω_p is the angular frequency, or PRF. From (16) we find that the mean *pulse* power is

$$R_{LF}(0)=\langle s_v(t_0')^2\rangle_{Av}=\eta^2\langle s_F(t_0')^4\rangle_{Av}, \quad (27)$$

and for the signal output we have

$$s_v(t_0')=\eta s_F^2(t_0'). \quad (28)$$

[12] The various forms of $R_{LF}(t)$ appear to have been derived independently by Fränz, Zeits. f. Hochfrequenztechnik, p. 140 (1941), who gave essentially the series expression, by D. O. North, who mentions it in a synopsis of a paper entitled "The Modification of Noise by Certain Non-Linear Devices," presented at the Winter Technical Meeting of the I.R.E., Jan. 28, 1944, and by G. E. Uhlenbeck in terms of the complete elliptic integrals E and K, in (M.I.T.) Radiation Laboratory Report 453, "Theory of Random Processes," Oct. 15, 1943. These results may also be obtained from Rice's expression (4.7-6), reference 8, for the correlation function of the noise output after half-wave linear rectification, with the help of (12) and the expansion of $(\cos\omega_0 t)^{2n}$.

In half-wave quadratic rectification we need merely divide (27) and (28) by 4. For video responses, then, (28) is simply

$$s_v = \eta s_F^2, \tag{29}$$

where s_v and s_F are the maximum values of $s_v(t_0')$ and $s_F(t_0')$, respectively. Hitherto we have denoted the time argument of the signal amplitude by t_0', distinguishing it from the argument t_0 of the carrier oscillations so that in integrating over the latter we could treat t_0' as constant. Since this integration has been performed, the prime can now be dropped.

To obtain the audio-spectrum we must take the Fourier components of the modulation. Now in aural reception we listen to a single component of the pulse, corresponding to the PRF or a higher harmonic thereof. Replacing t_0' by t in (28) and expanding $s_v(t)$ in a Fourier series, we find

$$s_v(t) = \sum_{-\infty}^{\infty} a_k \exp(ik\omega_p t), \quad \omega_p = 2\pi f_p = 2\pi/T, \tag{30}$$

where T is the period of the pulses and the amplitude of the kth harmonic of the PRF f_p is given by

$$a_k = \left\{ \frac{1}{T} \int_{-\infty}^{\infty} s_v(t) e^{-i\omega t} dt \right\}_{\omega = k\omega_p}, \tag{31}$$

since successive pulses are assumed not to overlap appreciably. Equation (30) may also be written as

$$s_v(t) = B_0 + \sum_{k=1}^{\infty} B_k \cos(k\omega_p t - \epsilon_{kp}), \tag{32}$$

where ϵ_{kp} is an uninteresting phase factor and

$$\begin{aligned} B_0 &= \left| \frac{1}{T} \int_{-\infty}^{\infty} s_v(t) e^{-i\omega t} dt \right|_{\omega=0}; \\ B_k &= \left| \frac{2}{T} \int_{-\infty}^{\infty} s_v(t) e^{-i\omega t} dt \right|_{\omega = k\omega_p}. \end{aligned} \tag{33}$$

If $\omega_k = \omega_p k$ remains small compared with the IF band width, we may set $\omega = 0$ in (33) because then the trigonometric factor varies but little during the time the pulse is on, and we may take $\sqrt{2}B_0$ as the r.m.s. amplitude s_a of the kth component we are listening to. With the help of (28), Eq. (33) becomes then

$$B_k \equiv s_a = \eta \frac{\sqrt{2}}{T} \int_{-\infty}^{\infty} s_F(t)^2 dt. \tag{34}$$

(d). *Signal Output of a Linear Detector in the Presence of Noise: Visual and Aural Reception*

The expression for the shape of the signal wave transmitted by a linear rectifier is complicated, because it is quite different from what it would be in the absence of noise, the interference from the latter giving rise to the phenomenon which is known as "modulation suppression." This phenomenon results in a suppression of the signal, or of the noise, whichever is small compared with the other. The detailed mathematical theory shows that the low frequency part of the rectified signal is given by[1,13–15]

$$s_v(t_0') = \frac{\beta n_F}{(2\pi)^{\frac{1}{2}}} {}_1F_1(-\tfrac{1}{2}; 1; -s_F(t_0')^2/n_F^2), \tag{35}$$

where ${}_1F_1$ is a confluent hypergeometric function and β, as before, is the dynamic transconductance of the rectifier. The function ${}_1F_1$ is specified in general by the series

$${}_1F_1(\alpha; \beta; x) = 1 + \frac{\alpha x}{\beta 1!} + \frac{\alpha(\alpha+1)x^2}{\beta(\beta+1)2!} + \cdots + \frac{\alpha_n x^n}{\beta_n n!} + \cdots. \tag{36}$$

The video response $s_v(t_0')$ represents the mean area under a cycle of the IF carrier, times the envelope of the modulation. In comparing signal and noise we must be careful to define what we mean by a signal and by noise, and to measure them in the same way. Hence, since noise is measured as an r.m.s. voltage, we must determine the signal similarly. This is the reason for taking the mean area under a cycle of the IF carrier as the signal criterion, rather than the envelope, a point which has not always been brought out in previous treatments of the subject.

Equation (35) as it stands includes both signal and d.c. noise; to obtain the increment in video

[13] See S. O. Rice, reference 8, Eq. (4.2–3).
[14] S. A. Goudsmit, "Comparison Between Signal and Noise" (M.I.T. Radiation Laboratory, Report No. 193, January 29, 1943).
[15] W. H. Jordan, "Action of Linear Detector on Signals in the Pressure of Noise" (M.I.T. Radiation Laboratory, Report No. 305, July 6, 1943).

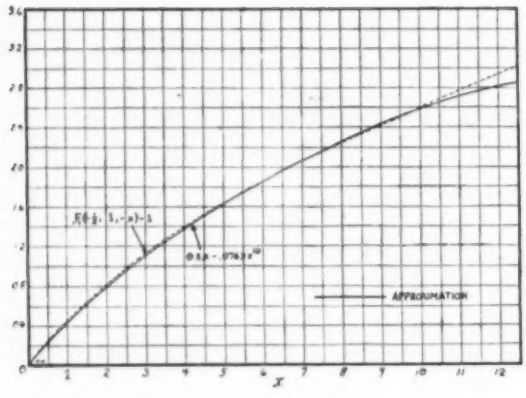


FIG. 9. Approximation of ${}_1F_1(-\frac{1}{2};1;-x)-1$ by $\alpha x-\gamma x^k$, $\alpha=\frac{1}{2}$, $\gamma=0.0763$, $k=1.50$.

response due to the signal we must subtract the mean d.c. response in the absence of the signal, thus obtaining

$$s_v(t_0')=\frac{\beta n_F}{(2\pi)^{\frac{1}{2}}}\times\{{}_1F_1(-\tfrac{1}{2};1;-s_F(t_0')^2/n_F^2)-1\},\quad(37)$$

inasmuch as ${}_1F_1(\alpha;\beta;0)=1$. At the instant of maximum signal, which is important in visual detection, we have by definition $s_v(t_0')=s_v$, $s_F(t_0')=s_F$, and Eq. (37) becomes

$$s_v=\frac{\beta n_F}{(2\pi)^{\frac{1}{2}}}\{{}_1F_1(-\tfrac{1}{2};1;-s_F^2/n_F^2)-1\},\quad(38)$$

where as before s_F is the maximum value of $s_F(t')$. When the signal is weak compared with the noise we find that (38) is

$$s_v\doteq\frac{\beta}{2(2\pi)^{\frac{1}{2}}}n_F\left(\frac{s_F}{n_F}\right)^2,\quad(s_F/n_F)^2\ll1,\quad(39)$$

which is a condition often encountered in practise. On the other hand, when the signal is very large compared to the noise, or what is the same thing, for a finite signal and negligible noise, we have with the help of the asymptotic expansion of ${}_1F_1$

$$s_v\doteq\frac{\beta\sqrt{2}}{\pi}s_F,\quad(s_F/n_F)^2\gg1,\quad(40)$$

which, as one would expect, is $1/\pi$ times the envelope of the IF wave.

In aural reception we listen to a single component of (37). With the aid of (30)–(34) we may write for the r.m.s. value s_a of this component

$$s_a=\frac{\beta n_F}{T\pi^{\frac{1}{2}}}\int_{-\infty}^{\infty}\{{}_1F_1(-\tfrac{1}{2};1;-s_F(t)^2/n_F^2)-1\}dt.\quad(41)$$

Although the power series for ${}_1F_1(-\frac{1}{2};1;-x)$ is absolutely convergent for all values of x, it does not converge rapidly enough for computation when x is large, say $x>1.5$. This difficulty may be overcome by approximating in a purely *ad hoc* fashion the integrand of (41) with a function of the form

$$F(x)=\alpha x-\gamma x^k.\quad(42)$$

Here α is the derivative of ${}_1F_1$ at $x=0$, and γ and k are determined from ${}_1F_1$ and its first derivative at some point $x=x_0$ which represents the maximum value of $s_F(t)^2/n_F^2$. We have then

where

$$\begin{aligned}\alpha&={}_1F_1'(0)=\tfrac{1}{2},\\ F'(x_0)&=\tfrac{1}{2}{}_1F_1(\tfrac{1}{2};2;-x_0),\\ F(x_0)&={}_1F_1(-\tfrac{1}{2};1;-x_0)-1,\\ \gamma&=[\alpha x_0-F(x_0)]x_0{}^k,\\ k&=\left(\frac{F'(x_0)-\alpha}{F(x_0)-\alpha x_0}\right)x_0.\end{aligned}\quad(43)$$

The necessary values of the hypergeometric functions may be obtained from Jahnke and Emde.[16] The approximation (42) requires that for each value of x_0 we calculate anew γ and k; however, in practise the error is not great if we select some value of x_0, say 6, and use the γ and k obtained in this case for all x_0 in the range $0\leq x_0\leq10$. Figure 9 compares our approximation with the correct curve. The difference in the region $0\leq x_0\leq10$ is seen to be quite small. Thus for subsequent work we choose $x_0=6$ and with this value find from (43) that

$$\gamma=0.0763;\quad k=1.50;\quad x_0=6.\quad(44)$$

Then the r.m.s. audio-component (41) after linear rectification in the presence of noise becomes finally:

$$s_a=\frac{\beta n_F}{2\pi^{\frac{1}{2}}T}\int_{-\infty}^{\infty}\{s_F(t)^2/n_F{}^2-\gamma s_F(t)^{2k}/n_F{}^{2k}\}dt.\quad(45)$$

[16] Jahnke and Emde (Stechert, New York, 1938), pp. 275 *et. seq.*

(e) *Comparison of Signal with Noise in Visual and Aural Reception*

When we listen or look for a signal with the help of our receiving apparatus, we do so against a noise background, man-made or natural, which tends to confuse or obscure our perception. Our ability to detect a signal will then obviously depend on the magnitude of the ratio of the signal at the point of reception (i.e., on the screen, at the earphones, or on a meter), to the appropriate noise background. In Sections (a)–(d) the various signal and noise voltages that apply in our problem have been determined; it remains now to compare them.

For visual reception involving quadratic detection we have from (18) and (29)

$$[s_v/n_v]_{\text{quad}}=(s_F/n_F)^2. \tag{46}$$

For linear rectification this ratio is modified to

$$[s_v/n_v]_{\text{lin}}=\rho_0[(s_F/n_F)^2-2\gamma(s_F/n_F)^{2k}],$$
$$\rho_0=\tfrac{1}{2}(4/\pi-1)^{-\frac{1}{2}}=0.957, \tag{47}$$

with the aid of (25), (38), and (42). *When the signal is small in comparison with the r.m.s. noise*, i.e., *when* $(s_F/n_F)^2\ll 1$, *we may neglect the second term in* (47), *and then the linear rectifier behaves almost exactly like the quadratic.* In fact, for most practical purposes we may set ρ_0 equal to unity, so that then the performance of both types of detector is identical. For strong signals, $s_F/n_F\geq 1$, the quadratic rectifier becomes noticeably superior, i.e., (>2 db), to the linear one, as shown by curve (a) of Fig. 2, and as described more fully in Sec. III following.

In aural work we find from (20), (23), and (34)

$$(s_a/n_a)_{\text{quad}}=\frac{1}{T(2\xi_0\Delta f_e)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\frac{s_F(t)^2}{n_F{}^2}dt, \tag{48a}$$

with

$$\xi_0=\int_0^{\infty}r_0(t)^2dt. \tag{48b}$$

The corresponding expression for the linear rectifier becomes, from (23), (26), and (45):

$$(s_a/n_a)_{\text{lin}}=\frac{\rho_0{}'}{T(2\Delta f_e)^{\frac{1}{2}}}$$
$$\times\int_{-\infty}^{\infty}\{s_F(t)^2/n_F{}^2-2\gamma(s_F(t)/n_F)^{2k}\}dt, \tag{49a}$$

where now

$$\rho_0{}'=\left[\sum_{n=0}^{\infty}\left(\frac{(2n)!}{2^{2n}n!(n+1)!}\right)^2\int_0^{\infty}r_0(t)^{2n+2}dt\right]^{-\frac{1}{2}}. \tag{49b}$$

Again if the signal-to-noise ratio is small we may omit the second term in the numerator of (49a). Furthermore, there is no serious error in discontinuing the infinite series in (49b) after the first term. Then we have $\rho_0{}'=\xi_0^{-\frac{1}{2}}$, with the result that (49a) and (48a) become identical. In a similar approximation we can replace the series involved in (24) and (25) by their first terms, and then the constant of proportionality ρ_0 in (47) becomes $\xi_0^{-\frac{1}{2}}$ instead of $0.957\xi_0^{-\frac{1}{2}}$, so that there is a like identity of results for the two detectors in the visual case.

When the series in (49b) is not replaced by its first term, the value of $\rho_0{}'$ depends slightly on the filter characteristics, being somewhere between $0.957\xi_0^{-\frac{1}{2}}$ and $\xi_0^{-\frac{1}{2}}$. In the Gaussian case the value of $\rho_0{}'$ is $0.973\xi_0^{-\frac{1}{2}}$. For practical purposes the difference in behavior of the two detectors is important only when we must consider the second term of the integrand of (49a). This makes the linear detector again less effective than the quadratic for strong signals (cf. curve (b) of Fig. 2). For the specific case of a Gaussian pulse and filter, consult Sec. IV, Part II.

III. Proof of the Fourier Transform Criterion

(a). *Visual Reception*

When the signal is detected visually, the best filter is by definition, according to our "simple" standards (see Sec. I, Part I), that which gives the largest value of the ratio s_v/n_v of peak video signal to r.m.s. video noise. This is equivalent to demanding that the ratio s_F/n_F of the maximum signal amplitude to r.m.s. noise amplitude after filtering, but before rectification, be as large as possible. By the maximum signal amplitude we mean the greatest absolute value of the "modulation factor" $s_F(t)$ for the signal $\sqrt{2}s_F(t)\cos\omega_0 t$ which is passed by the filter. The quantities s_v, n_v are by no means the same as s_F, n_F, respectively. For instance, the mean video noise power $n_v{}^2$ is but a fraction of the total noise $n_F{}^2$ passing through the filter, with the fraction depending on the type of detector, but not on the filter

design (cf. especially Eqs. (18) and (25)). However, the preceding analysis has shown that s_v/n_v is a monotonically increasing function of s_F/n_F. Namely, Eq. (46) shows that for the quadratic rectifier s_v/n_v is proportional to s_F^2/n_F^2, as might have been guessed. For the linear detector s_v/n_v is proportional to s_F^2/n_F^2 only if $s_F^2/n_F^2 \ll 1$, but the hypergeometric function ${}_1F_1(-\frac{1}{2}; 1; -s_F^2/n_F^2)$ possesses the monatonic property for the entire domain in which we are interested. Hence maximization of s_v/n_v also implies maximization of s_F/n_F.

It was shown in Section 1, Part II, that $s_v(t)$ can be computed, as we might suspect, by forgetting the carrier entirely, and regarding the origin of filter response as at the carrier rather than at zero absolute frequency. Thus if the signal entering the (IF) filter is $\sqrt{2}s_I(t)\cos\omega_0 t$, the value of the amplitude function $s_F(t)$ after filtering is, from (7b), at any given time t_0

$$s_F(t_0) = \frac{1}{\sqrt{2}}\left|\frac{1}{\pi}\int_{-\infty}^{\infty} G_{IF}(\omega)S_I(\omega)\exp(i\omega t_0)d\omega\right|. \quad (50)$$

The noise energy passed by the filter is given by (11) when t is set equal to zero. The input energy per pulse is

$$E = \int_{-\infty}^{\infty} s_I(t)^2 dt = \frac{1}{\pi}\int_{-\infty}^{\infty} |S_I(\omega)|^2 d\omega, \quad (51)$$

the second relation following in virtue of Plancherel's theorem[17] and Eq. (5). Since (51) is independent of filter design the ratio which we desire to maximize can be regarded equally well as $(s_F/n_F)^2$ or s_F^2/n_F^2E. With the latter form, except for constants which do not depend on pulse or meter shape, the ratio becomes

$$\frac{\left|\int_{-\infty}^{\infty} G_{IF}(\omega)S_I(\omega)\exp(i\omega t_0)d\omega\right|^2}{\int_{-\infty}^{\infty}|S_I(\omega)|^2 d\omega \int_{-\infty}^{\infty}|G_{IF}(\omega)|^2 d\omega}. \quad (52)$$

By the Schwartz inequality, this expression has the maximum value unity, and this is achieved when

$$G_{IF}(\omega) = [S_I(\omega)\exp(i\omega t_0)]^* = S_I(\omega)^*\exp(-i\omega t_0). \quad (53)$$

Equation (53) shows that except for a phase factor *the filter characteristic should be the conjugate of the Fourier transform of the pulse*, a condition we have quoted many times. The distinction between $S_I'(\omega)$ and $S_I(\omega)^*$ disappears for a symmetrical pulse, as then $S_I(\omega)$ is real. Otherwise one must match the filter to the conjugate of the Fourier transform rather than to the transform itself.

The theorem just given has already been derived, cf. North.[1] A different method of proof is usually employed, *viz.*, the calculus of variations. The possibility of using the Schwartz inequality, as we do here, has also been noted by Dr. Henry Wallman,[18] and is contained as a special case of a more general theory due to Wiener. More important, in our opinion, than the mathematical nature of the proof, is the analysis and discussion given in the first paragraph of the present section, which shows that maximization of s_v/n_v also implies maximization of s_F/n_F. The usual proofs consider only the extremum of s_F/n_F, which mathematically does not *a priori* correspond to an extremum in s_v/n_v, though it is physically very reasonable that this should be the case. It is particularly to be emphasized that the maximum achievable value of the ratio (52) is quite independent of the shape of the noise spectrum and of the pulse $s_I(t)$. This means that with perfect matching the performance with visual presentation is independent of pulse shape, a result already quoted in Part I. In other words, *with our "simple" criterion*, the only thing that counts at match is just the input energy per pulse.

The phase factor $\exp(i\omega t_0)$ requires some discussion. It is determined by the time at which the signal is a maximum, though the magnitude of this maximum is unaffected by the choice of t_0. At first sight it might seem that the time at which the signal reaches its maximum can be chosen at will by giving the phase factor $\exp(i\omega t_0)$ of the filter the proper value. However, this is

[17] See, for example, E. C. Kemble, *Fundamental Principles of Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1935), p. 36.

[18] "Realizability of Filters" (M.I.T. Radiation Laboratory Report No. 637, Dec. 8, 1944).

not the case, for the filter must be capable of actual construction. On obvious physical grounds it is impossible to construct a filter which will give the response of a signal at the output before this signal even reaches the input. A filter which does not have this absurd characteristic we shall call "physically realizable." Mathematically, this can be shown equivalent to demanding that $G_{IF}(\omega)$ have no singularities in the lower half of the complex plane. Any polynomial structure of the type $\Pi_j(\omega-i\delta_j)^{-1}$, $\delta_j>0$ is clearly physically realizable and corresponds to a series of cascaded, single-tuned circuits. The gaussian structure which we study rather explicitly later in the report is not physically realizable, but as shown by Wallman,[18] this objection is rather academic. Namely, the gaussian behavior can be regarded as the limit of cascading a very large number of circuits. In the limit the time delay would have to be infinite, but the gaussian behavior, as far as frequency dependence is concerned, does not differ materially for practical purposes from the cascading of several circuits with only a finite time-lag. So although it is not possible to build a strictly gaussian system, one can find an actual system whose filtering action is very nearly equivalent to it, and, in fact, practical filters constructed by cascading will have an essentially gaussian response curve, to all intents and purposes.

(b) *Aural Reception*

We must now examine whether the Fourier transform criterion also applies to aural as well as to visual reception. Here the proof is not quite so immediate. First let us consider the square detector. As we are interested only in one audio-component, the effective signal-to-noise ratio, except for a constant factor A, is

$$(s_a/n_a)^2=A\left(\int_{-\infty}^{\infty}s_F(t)^2dt\right)^2\Big/\int_{-\infty}^{\infty}r_0(t)^2dt, \quad (54)$$

where $r_0(t)$ is given by (12). The denominator of (54) is obtained directly from (17), and the numerator follows at once from (34). Since the Fourier transform of $s_F(t)$ is $S_I(\omega)G_{IF}(\omega)\sqrt{2}$, (*vide* Eq. (7b)), and that of $r_0(t)$ is proportional to $|G_{IF}(\omega)|^2$, by Plancherel's theorem (54) can be written

$$(s_a/n_a)^2=B\left(\int_{-\infty}^{\infty}|G_{IF}(\omega)|^2|S_I(\omega)|^2d\omega\right)^2\Big/\int_{-\infty}^{\infty}|G_{IF}(\omega)|^4d\omega. \quad (55)$$

If we divide by $\int_{-\infty}^{\infty}|S_I(\omega)|^4d\omega$, which is constant with respect to any given type of pulse, we have a ratio to which we can apply the Schwarz inequality. It follows immediately that for a maximum signal-to-noise ratio we must have

$$|S_I(\omega)|^2=C|G_{IF}(\omega)|^2, \quad (56)$$

when C is an unimportant constant which we may set equal to unity without loss of generality.

Two remarks particularly should be made in connection with (56). One is that the question of the phase of the filter response does not arise, for the only quantity specified is $|G_{IF}(\omega)|$, and the phase shift can be an arbitrary function of frequency. In the visual case, on the other hand, the phase had to be the same as that of the conjugate of the Fourier transform of the signal, except for an arbitrary linear term determined by the time at which the signal was a maximum. Such a time obviously cannot enter in connection with audio detection, as here attention is focused on one particular component rather than on the maxima and minima obtained by superposing all the components.

The other remark is that with audio-reception the response at match is conditioned by shape rather than by just energy per pulse. If the latter alone were to enter, (55) should be divided by a factor $\int_{-\infty}^{\infty}|S_I(\omega)|^2d\omega$, rather than $\int_{-\infty}^{\infty}|S_I(\omega)|^4d\omega$. The resulting increased effectiveness of long pulses has already been discussed in Section V, Part I.

The proof that the Fourier transform criterion gives the best results in aural detection cannot be given in as clear-cut form for the linear as for the square-law detector. If we assume the signal is weak, so that we may use (39) and neglect all terms beyond $r_0^2(t)$ in (26), the ratio to be maximized is the same as for the square

detector and our previous proof will apply. Numerical examination of particular cases shows that the neglected terms do not disturb the location of the point of maximum efficiency. We may remark, moreover, that if it were not for the phenomenon of modulation suppression, the width of the IF filter would be a matter of indifference in aural detection. In actual practice, the true relationship is, so to speak, a compromise between (39) and (40), being much closer to the former than to the latter. Since in (40) the filter design does not enter, and since with (39) the Fourier transform criterion is valid, we can see qualitatively that in any intermediate case, use of this criterion will presumably bring the best results.

It can be shown, from the same sort of analysis, that the introduction of a video-filter at match gives no improvement of the signal to noise ratio. We will not give the details here, as the problem is amply covered elsewhere.[19]

IV. Explicit Calculation for a Gaussian Pulse and Filter

The general expression for the aural and visual signal-to-noise ratios have been obtained in Section II, Part II, but before we can arrive at complete numerical results on all aspects of the problem, we must specify the pulse $s_I(t)$ entering the IF and the response $G_{IF}(\omega)$ of the IF filter. Let the unfiltered envelope be

$$\sqrt{2}s_I(t)=S_0\exp(-\omega_g^2t^2/2),\quad \omega_g=2\pi f_g, \tag{57}$$

where S_0 is the maximum amplitude of the disturbance, and ω_g is an angular frequency determining the width of the pulse. With the help of (5), the amplitude spectral distribution becomes

$$S_I(\omega)=(\pi/2)^{\frac{1}{2}}\omega_g^{-1}S_0\exp(-\omega^2/2\omega_g^2). \tag{58}$$

Now for the (shifted) *amplitude* response of the IF filter we have

$$G_{IF}(\omega)=\exp(-\omega^2/2\omega_b^2),\qquad \omega_b=2\pi f_b,\quad(\omega=\omega'-\omega_0), \tag{59}$$

where ω_b similarly determines the filter width.

[19] J. W. Lawson and G. E. Uhlenbeck, Radiation Laboratory Series, book 24.

In terms of the equivalent rectangular filter discussed in Section II, Part II, from Eq. (22) we find for the pulse and IF filter

$$\Delta f_{IF}=(\Delta f_b)=\pi^{\frac{1}{2}}f_b;\qquad \Delta f_I=(\Delta f_g)=(1/\Delta T_g)=\pi^{\frac{1}{2}}f_g. \tag{60}$$

With the aid of (7b) we see from (58) and (59) that the filtered wave becomes

$$\sqrt{2}s_F(t)=(bS_0/\sqrt{2}\omega_g)\exp(-t^2b^2/4)=\sqrt{2}s_F\exp(-t^2b^2/4), \tag{61a}$$

where

$$b\equiv\sqrt{2}\omega_b\omega_g/(\omega_g^2+\omega_b^2)^{\frac{1}{2}};\quad s_F\equiv bS_0/\sqrt{2}\omega_g. \tag{61b}$$

From (51) and (57) the input energy in a single gaussian pulse entering the IF is

$$E=\int_{-\infty}^{\infty}s_I(t)^2dt=S_0^2\pi^{\frac{1}{2}}/2\omega_g, \tag{62}$$

and since the input power to the IF is (energy/pulse)$\times$(no. of pulses), we have

$$P_{\text{in}}=E/T=S_0^2\pi^{\frac{1}{2}}/2\omega_gT, \tag{63}$$

where T is the period of the pulses. From (8) and (59) the mean square noise amplitude passed by the gaussian IF is

$$n_F^2=\frac{W_0}{2\pi}\int_{-\infty}^{\infty}\exp(-\omega^2/\omega_b^2)d\omega=W_0\omega_b/2\pi^{\frac{1}{2}}, \tag{64}$$

so that from (61a, b) we find that

$$s_F(t)^2/n_F^2=(s_F^2/n_F^2)\exp(-t^2b^2/2)=\frac{\pi^{\frac{1}{2}}b^2S_0^2}{2W_0\omega_b\omega_g^2}\exp(-t^2b^2/2). \tag{65}$$

Now let us define a quantity σ_0 by

$$\sigma_0\equiv(s_F/n_F)_{\text{at match}}, \tag{66}$$

where σ_0 is a value of s_F/n_F, before rectification, which is obtained when a gaussian pulse, whose pulse repetition frequency is T_0^{-1}, enters an IF filter matched to the pulse, i.e., $\omega_b=\omega_g$. Let P_0 represent the input power to the IF under these circumstances. Then from (59) and (63) we obtain

$$\sigma_0^2=\pi^{\frac{1}{2}}S_0^2/2W_0\omega_b,$$

and

$$P_0=S_0^2\pi^{\frac{1}{2}}/2\omega_bT_0=W_0\sigma_0^2/T_0, \tag{67}$$

which, applied to (65) gives us finally

$$s_F(t)^2/n_F^2=(s_F^2/n_F^2)\exp(-t^2b^2/2)$$

$$=\left(\frac{2\omega_g\omega_b}{\omega_g^2+\omega_b^2}\right)\sigma_0^2\left(\frac{P_{\text{in}}}{P_0}\frac{T}{T_0}\right)\exp(-t^2b^2/2). \quad (68)$$

To examine the effects of variable pulse and filter widths on the possibly perceptible signal-to-noise ratios, we must then distinguish two cases:

Case I: Here the filter width ($\sim\omega_b$) remains constant, while the pulse width ($\sim\omega_g$) may have different values. Let

$$\bar{\lambda}\equiv\omega_g/\omega_b=f_g/f_b,$$

and

$$\therefore\quad b=\sqrt{2}\omega_g(1+\bar{\lambda}^2)^{-\frac{1}{2}}. \quad (69)$$

From (58) it follows that

$$(s_F/n_F)^2=\frac{2\bar{\lambda}}{1+\bar{\lambda}^2}\sigma_0^2\bar{P}\bar{T}, \quad (70)$$

where $\bar{P}\equiv P_{\text{in}}/P_0$, and $\bar{T}=T/T_0$, a normalization made in terms of quantities experimentally determined when the pulse and filter are matched. In a similar fashion, we find for *Case II*, where the pulse width ($\sim\omega_g$) is kept fast and the filter width ($\sim\omega_b$) is varied

$$\lambda\equiv\omega_b/\omega_g(=\bar{\lambda}^{-1}),$$

and

$$\therefore\quad b=\lambda\sqrt{2}\omega_g(1+\lambda^2)^{-\frac{1}{2}}, \quad (71)$$

and, like (70)

$$(s_F/n_F)^2=\frac{2\lambda}{1+\lambda^2}\sigma_0^2\bar{P}\bar{T}. \quad (72)$$

When (70) and (72) are used in calculations involving the linear rectifier we must be careful that $(s_F/n_F)^2$ does not exceed 9 or 10, in accordance with the approximation (42). See also Fig. 9.

(*a*). *Visual Reception with a Quadratic or Linear Rectifier*

For visual reception we may write at once from (46), (70), and (72) that

$$(s_v/n_v)_{\text{quad}}=\frac{2\Lambda}{1+\Lambda^2}\sigma_0^2\bar{P}\bar{T}, \quad (73)$$

where $\Lambda=\lambda$ or $\bar{\lambda}$, as the case may be. Now when $(s_v/n_v)_{\text{quad}}$ represents the minimum detectable signal-to-noise ratio, $(s_v/n_v)/(s_v/n_v)_{\text{at match}}$ must equal unity, since the minimum perceivable signal is by definition a constant σ_0^2, from Eq. (73). Then it is evident that

$$1\equiv(s_v/n_v)_{\text{quad}}/(s_v/n_v)_{\text{at match }\Lambda=1}=\frac{2\Lambda}{1+\Lambda^2}\bar{P}\bar{T}, \quad (74)$$

a somewhat special case of a more general result obtained by M. C. Wang and G. E. Uhlenbeck at (M.I.T.) Radiation Laboratory, and considered by them elsewhere.

For visual reception following a linear detector we may similarly write, from (47), (70), and (72),

$$(s_v/n_v)_{\text{lin}}=\rho_0\left(\frac{2\Lambda}{1+\Lambda^2}\right)\bar{P}\bar{T}\sigma_0^2$$

$$\times\left\{1-2\gamma\left(\frac{2\Lambda\sigma_0^2\bar{P}\bar{T}}{1+\Lambda^2}\right)^{k-1}\right\}, \quad \begin{matrix}\rho_0=0.957,\\ \gamma=0.0763,\\ k=1.50.\end{matrix} \quad (75)$$

Analogous to (74) we have

$$1\equiv(s_v/n_v)_{\text{lin}}/(s_v/n_v)_{\text{at match }\Lambda=1}$$

$$=\frac{2\Lambda}{1+\Lambda^2}\bar{P}\bar{T}\left(\frac{1-2\gamma\{2\Lambda\sigma_0^2\bar{P}\bar{T}/(1+\Lambda^2)\}^{k-1}}{1-2\gamma\sigma_0^{2k-2}}\right). \quad (76)$$

We see at once that if $(s_F/n_F)^2\ll1$, we may neglect the second terms of (75) and (76), so that both of these equations become identical with (73) and (74), for the square-law rectifier, on the slight additional assumption that we treat $\rho_0=1$.

(*b*). *Aural Reception with a Quadratic and Linear Rectifier*

For aural reception we must employ (48a, b) and (49a, b). The normalized correlation $r_0(t)$ is obtained for the gaussian case from (13), and is $r_0(t)=\exp(-t^2\omega_b^2/4)$, whereupon ξ_0, cf. Eq. (48b), becomes $(\pi/2\omega_b^2)^{\frac{1}{2}}$. Finally, with the aid of (48a, b), (65), (70), and (72), and following the procedure of Section (a) above we obtain

$$(s_a/n_a)_{\text{quad}}/(s_a/n_a)_{\text{at match }\bar{\lambda}=1}$$

$$=[2/(1+\bar{\lambda}^2)]^{\frac{1}{2}}\bar{P},\ \text{Case I}, \quad (77a)$$

and for Case II

$$(s_a/n_a)_{\text{quad}}/(s_a/n_a)_{\text{at match }\lambda=1}$$

$$=[2\lambda/(1+\lambda^2)]^{\frac{1}{2}}\bar{P},\ \text{Case II}. \quad (77b)$$

Equations (77a) and (77b) are discussed at some length in Section V, Part I. In a similar fashion we obtain from (49a, b), (65), (70), (77) and Section (a) the ratios corresponding to the linear detector:

$$\frac{(s_a/n_a)_{\text{lin}}}{(s_a/n_a)_{\text{at match}}} \equiv 1 = \frac{[2/(1+\bar{\lambda}^2)]^{\frac{1}{2}}\bar{P}(1-2\gamma k^{-\frac{1}{2}}\{[2\bar{\lambda}/(1+\bar{\lambda}^2)]\sigma_0^2\bar{P}\bar{T}\}^{k-1})}{1-2\gamma k^{-\frac{1}{2}}\sigma_0^{2k-2}}, \quad \text{Case I,} \tag{78a}$$

and

$$\frac{(s_a/n_a)_{\text{lin}}}{(s_a/n_a)_{\text{at match}}} \equiv 1 = \frac{[2\lambda/(1+\lambda^2)]^{\frac{1}{2}}\bar{P}(1-2\gamma k^{-\frac{1}{2}}\{[2\lambda/(1+\lambda^2)]\sigma_0^2\bar{P}\bar{T}\}^{k-1})}{1-2\gamma k^{-\frac{1}{2}}\sigma_0^{2k-2}}, \quad \text{Case II.} \tag{78b}$$

These equations are also discussed in some detail in Sections V and VI, Part I, and are illustrated in Figs. 4 and 5.

The above aural signal-to-noise ratios have all been computed on the assumption that there is no gating of the noise before rectification. If a gate is in operation, usually before the IF or rectifier stages, the background audio-noise voltage is simply reduced by the square-root of the gating factor $G \equiv \tau_G/T$, where τ_G is the duration of the "on-" period, or transmitting interval of the gate, i.e., the time interval during which noise may enter the IF or rectifier. This is a simple multiplicative factor, provided the duration of the "on-" period is sufficiently long to permit the reception of consecutive signals from the particular target under scrutiny. A more thorough discussion of gating is given in Section VI, Part I. We may say, then, that for either type of rectifier the gated background audio-noise voltage (r.m.s.) becomes

$$n_{aG} = G^{\frac{1}{2}}n_a = (\tau_G/T)^{\frac{1}{2}}n_a. \tag{79}$$

When the effect of (79) is considered in deriving the results of (77a) and (77b), we observe that these latter are modified to

$$1 = \left(\frac{2}{1+\bar{\lambda}^2}\right)^{\frac{1}{2}}\bar{P}\bar{T}^{\frac{1}{2}}, \ \text{(I)},$$

and

$$1 = \left(\frac{2\lambda}{1+\lambda^2}\right)^{\frac{1}{2}}\cdot\bar{P}\bar{T}^{\frac{1}{2}}, \ \text{(II)}. \tag{80}$$

A similar modification takes place in (78a) and (78b), the right members of which are now multiplied by $\bar{T}^{\frac{1}{2}}$. Curves of (78a) and (78b) for gating are shown in Figs. 4 and 5, for the condition of match.

(c). *Comparison of the Aural and Visual Behavior of the Linear and Quadratic Detectors*

From the results of the preceding pages we are in a position to compare conveniently the visual performance of the two types of rectifier, and to make the comparison in the aural case, as well. Let us examine first, briefly, the ratio between the visual linear response and the corresponding one for the quadratic. From (73) and (75) we obtain at once

$$(s_v/n_v)_{\text{lin}}/(s_v/n_v)_{\text{quad}} = \rho_0\left\{1-2\gamma\left[\frac{2\Lambda}{1+\Lambda^2}\sigma_0^2\bar{P}\bar{T}\right]^{k-1}\right\} \tag{81}$$

which holds equally well for either Case I or Case II. At match $\bar{P}=\bar{T}=1$ and $\Lambda=1$, and we have

$$[(s_v/n_v)_{\text{lin}}/(s_v/n_v)_{\text{quad}}]_{\text{match}} = 0.957(1-0.1526\sigma_0), \quad \sigma_0 \leq 10^{\frac{1}{2}}. \tag{82}$$

For aural reception we obtain expressions analogous to (81) and (82) by the same process used in arriving at (77–78), namely

$$(s_a/n_a)_{\text{lin}}/(s_a/n_a)_{\text{quad}} = \mu_0\left(1-2\gamma k^{-\frac{1}{2}}\left\{\frac{2\Lambda}{1+\Lambda^2}\sigma_0^2\bar{P}\bar{T}\right\}^{k-1}\right), \tag{83}$$

where

$$\mu_0 = \left\{\sum_{n=0}^{\infty}[(2n)!/2^{2n}n!(n+1)!]^2\frac{1}{(n+1)^{\frac{1}{2}}}\right\}^{-\frac{1}{2}} = 0.973,$$

and where either f_b or f_g may be constant. Equation (83) becomes particularly simple at match ($\Lambda=\bar{P}=\bar{T}=1$), where $k=1.50$, yielding

$$\{(s_a/n_a)_{\text{lin}}/(s_v/n_v)_{\text{quad}}\}_{\text{at match}} = 0.973(1-0.1246\sigma_0), \quad \sigma_0 \leq 10^{\frac{1}{2}}, \tag{84}$$

which is not very different from (82). It might be pointed out that (82) is independent of pulse and filter shape, the only requirement being that the two are properly matched. This is very nearly true of (84), though not strictly so, as the factor $k^{-\frac{1}{2}}$ in (83) appears as a consequence of the gaussian structure of the filter and pulse. However, for all practical purposes we may consider the aural comparison also independent of pulse and filter shape, at match. Even when a condition of match does not obtain, if $(s_F/n_F)^2 \ll 1$, i.e., $2\Lambda\sigma_0^2\bar{P}\bar{T}/(1+\Lambda)^2 \ll 1$, which is the usual, practical situation, the performance of the two rectifiers is identical. Figure 2 demonstrates this, and shows how the behavior of the linear detector departs from that of quadratic. This departure in the customary region of operation does not exceed 1.5 db. Note that the ratios (83) and (84) hold equally well for gated noise, in the aural case.

V. General Theory for the Meter

Instead of listening to the fundamental (or a harmonic) of the PRF against the noise background, passed by a narrow audio-filter, we may rectify this component and its surrounding noise, and pass the mixture into a meter with a very narrow low pass filter, say from $\frac{1}{10}$ to 1 cycle wide. Our problem is to obtain an expression for the signal-to-noise ratio as recorded by the meter, i.e., to determine s_m/n_m and to examine how this ratio depends on such factors as audio- and meter filter widths, signal to noise in the audio-filter, etc.

Let us consider first the noise entering the meter. This results from the rectification of the noise passed by the narrow audio-filter centered about the harmonic of the PRF under examination. Because this filter is quite narrow, say 10–100 cycles, and conceivably less, we cannot always assume as we did for the case of aural detection, that the resulting very low frequency continuum, $W_{LF}(\omega)_a$, has a uniform spectral density, and can be replaced by $W_{LF}(0)_a$ in our derivation. We then write, following (21a), that the mean noise power in the meter is

$$n_m^2 = \frac{1}{2\pi}\int_0^\infty |G_m(\omega)|^2 W_{LF}(\omega)_a d\omega, \qquad (85)$$

where $G_m(\omega)$ is the amplitude response of the low pass meter filter. Although the low frequency noise is no longer gaussian after rectification by the second detector, the audio-filter is assumed sufficiently narrow so that the noise passed by it may be considered random, with its gaussian characteristics restored.[20] Then, to determine the low frequency spectrum after rectification we need only parallel the analysis of Section II, Part II, except that the signal, after being passed by the audio-filter, is now periodic, rather than highly pulsed. From (26) applied to the narrow portion of the *audio*-spectrum, instead of to the IF, we obtain for a linear rectifier in the meter circuit

$$W_{LF}(\omega)_a = \frac{n_a^2\beta_0^2}{2\pi}\sum_{n=0}^{\infty}\left\{\frac{(2n)!}{2^{2n}n!(n+1)!}\right\}^2 \times \int_0^\infty r_0(t)_a^{2n+2}\cos\omega t dt, \qquad (86)$$

where β_0 is the transconductance of the rectifier, and $r_0(t)_a$ is the normalized correlation function of the random noise passed by the audio-filter, which may easily be obtained from (12) and (13) on replacing $G_{IF}(\omega)$ by $G_a(\omega)$. Since the audio-filter is narrow compared to the PRF harmonic, we may safely extend the limits on the integration to $-\infty$ in determining $r_0(t)_a$, and for the mean audio-noise power we find

$$n_a^2 = \frac{W_{LF}(0)}{2\pi}\int_{-\infty}^{\infty}|G_a(\omega)|^2 d\omega, \qquad (87)$$

from (21b). As before, the argument of $G_a(\omega)$ is the frequency relative to an origin at the carrier, now the PRF or one of its harmonics. $W_{LF}(0)$, of course, depends on the characteristic of the second detector used after the IF. If the noise is gated, we may replace n_a by n_{aG} in subsequent formulae, remembering that the two forms are related by Eq. (79).

In case the rectification is quadratic rather than linear, we may write in place of (86), in virtue of (17) and (19),

$$W_{LF}(\omega)_a = 4\eta_0^2 n_a^4\int_0^\infty r_0(t)_a^2\cos\omega t dt, \qquad (88)$$

with η_0 the detector constant.

[20] See the first part of section 4.3, of Rice's article (reference 8).

The signal-to-noise ratio in the meter is obtained at once by analogy with (38) and (42) for the linear rectifier, and is

$$s_m/n_m=\frac{\beta_0}{2(2\pi)^{\frac{1}{2}}}\left(\frac{n_{a(G)}}{n_m}\right)\{(s_a/n_{a(G)})^2 - 2\gamma(s_a/n_{a(G)})^{2k}\}, \quad s_a/n_{a(G)}<10^{\frac{1}{2}}. \quad (89)$$

For the quadratic detector, Eq. (89) is modified by virtue of (29) to

$$s_m/n_m=\eta_0 s_a^2/n_m. \quad (90)$$

The expressions for $s_a/n_{a(G)}$ depend on the circuit preceding the meter. The proper forms are given in Section II, (e), Part II. With the above results we are now ready to proceed to specific problems.

VI. An Explicit Treatment for the Meter

Let us determine s_m/n_m for a meter following a linear rectifier. We assume further that the audio- and meter filters are essentially gaussian and have the responses

$$G_a(\omega)=\exp(-\omega^2/2\omega_a^2), \qquad G_m(\omega)=\exp(-\omega^2/2\omega_m^2), \quad (91)$$

where ω_a and ω_m are analogous to ω_b. The correlation $r_0(t)_a$ becomes with the help of (91)

$$r_0(t)_a=\exp(-t^2\omega_a^2/4). \quad (92)$$

With this result we may write for the low frequency spectrum after rectification, from (86),

$$W_{LF}(\omega)_a=\frac{n_a^2\beta_0^2}{2\omega_a(2\pi)^{\frac{1}{2}}}\times\sum_{n=0}^{\infty}\left\{\frac{(2n)!}{2^{2n}n!(n+1)!(n+1)^{\frac{1}{2}}}\right\}^2 \times\exp[-\omega^2/2\omega_a^2(n+1)]. \quad (93)$$

Now we find from (85) that the mean noise power is

$$n_m^2=\frac{n_a^2\beta_0^2}{8\pi\sqrt{2}}\sum_{n=0}^{\infty}\left\{\frac{(2n)!}{2^{2n}n!(n+1)!(n+1)^{\frac{1}{2}}}\right\}^2 \times\frac{(\omega_m/\omega_a)}{[1+(\omega_m/\omega_a)^2\frac{1}{2}(n+1)^{-1}]^{\frac{1}{2}}}. \quad (94)$$

Applying this directly to (89) we obtain for the signal-to-noise ratio in the meter

$$s_m/n_m=2^{\frac{1}{2}}(\textstyle\sum_n)^{-\frac{1}{2}} \times\{(s_a/n_{a(G)})^2-2\gamma(s_a/n_{a(G)})^{2k}\}, \quad (95)$$

where $\sum_n$ is the series in (94). Our result (95) may be simplified if we disregard the terms in $(\omega_m/\omega_a)^2[1/2(n+1)]$, $n\geq 1$, which appear in $\sum_n$, and which for our problem are quite small compared to the term in $n=0$. This approximation becomes progressively better the smaller the ratio ω_m/ω_a. Then the summation takes the value

$$\sum_n=\frac{1.059(\omega_m/\omega_a)}{[1+\frac{1}{2}[\omega_m/\omega_a)^2]^{\frac{1}{2}}}=\frac{1.059(\Delta f_m/\Delta f_a)}{[1+\frac{1}{2}(\Delta f_m/\Delta f_a)^2]^{\frac{1}{2}}}, \quad (96)$$

where Δf_m and Δf_a are the widths of the equivalent rectangular filters, *vide* Eq. (22). We have then for gated noise

$$\frac{s_m}{n_m}=1.156\frac{\{1+\frac{1}{2}(\Delta f_m/\Delta f_a)^2\}^{\frac{1}{2}}}{(\Delta f_m/\Delta f_a)^{\frac{1}{2}}} \times\left\{(s_a/n_{a(G)})^2-2\gamma\left(\frac{s_a}{n_{a(G)}}\right)^{2k}\right\}, \; s_a/n_{a(G)}\leq 10^{\frac{1}{2}}. \quad (97)$$

With this result we are at once able to compare, *for a given set of operating conditions*, the relative merits of the meter, the "ear-scope," or earphone detection, and the *A*-scope, provided we know in each case what the physically minimum values of s_m/n_m, $s_a/n_{a(G)}$, s_v/n_v are. Once these are known, the comparison follows immediately, as they are all functions of the single quantity s_F/n_F. (See Section VII, Part I.)

We are particularly interested in determining, for a given apparatus, how wide the audio-filter, Δf_e, must be in the "earscope" method, as compared with the filters used in connection with periodic meter detection. We will assume that the minimum detectable signal-to-noise ratios are related in either case by

$$s_m/n_m=A[s_a/n_{a(G)}]_{\text{earscope}}, \quad (98)$$

where the constant A is determined experimentally. One condition of comparison is then, from (98)

$$[s_a/n_{a(G)}]_{\text{earscope}}=\alpha_0/A, \quad (99)$$

where now α_0 is the value of the minimum detectable signal-to-noise ratio when a periodic metering scheme is used, into which $s_a/n_{a(G)}=\alpha$ has been introduced. The quantity α may be defined by the following relation

$$[s_a/n_{a(G)}]_{\text{earscope}}\equiv\alpha(\Delta f_a/\Delta f_e)^{\frac{1}{2}}$$

$$=\left[\frac{s_a}{n_{a(G)}}\right]\left(\frac{\Delta f_a}{\Delta f_e}\right)^{\frac{1}{2}}, \quad (100)$$

which is an obvious consequence of the fact that the noise power passed by the audio-filter is proportional to its width. Using (97)–(100), we find first that

$$\alpha=\alpha_0(\Delta f_e/\Delta f_m)^{\frac{1}{2}}/(\Delta f_a/\Delta f_m)^{\frac{1}{2}}. \quad (101)$$

Letting $y\equiv\Delta f_e/\Delta f_m$ and $x\equiv\Delta f_a/\Delta f_m$, we have with the help of (97) and (101)

$$\alpha_0=1.156(1+\tfrac{1}{2}x^{-2})^{\frac{1}{2}}x^{\frac{1}{2}}\left\{\left[\frac{\alpha_0}{A}(y/x)^{\frac{1}{2}}\right]^2\right.$$

$$\left.-2\gamma\left[\frac{\alpha_0}{A}(y/x)^{\frac{1}{2}}\right]^3\right\}, \quad \gamma=0.0763, \quad (102)$$

which reduces simply to

$$1=(y/x)^4(2x^2+1)(\alpha_0/A^2 1.059^{\frac{1}{2}})^4$$

$$\times\{1-2\gamma\alpha_0(y/x)^{\frac{1}{2}}/A\}. \quad (103)$$

Let

$$a_0\equiv\alpha_0/A^2(1.059)^{\frac{1}{2}}=0.971\alpha_0/A^2;$$
$$b_0\equiv 2\gamma\alpha_0/A=0.1526\alpha_0/A; \quad z\equiv y/x. \quad (104)$$

Then (103) becomes

$$(za_0)^4(2x^2+1)(1-b_0z^{\frac{1}{2}})^4=1, \quad (105)$$

subject to the requirement that z is always chosen so that $(1-b_0z^{\frac{1}{2}})za_0<1$. If the detector just preceding the meter is quadratic, instead of linear, Eq. (97) may be written, from (85), (88), (90):

$$\frac{s_m}{n_m}=2^{\frac{1}{4}}\frac{\{1+\frac{1}{2}(\Delta f_m/\Delta f_a)^2\}^{\frac{1}{4}}}{(\Delta f_m/\Delta f_a)^{\frac{1}{2}}}$$

$$\times\left[\frac{s_a}{n_{a(G)}}\right]^2; \quad 2^{\frac{1}{4}}=1.189. \quad (106)$$

This reduces simply to

$$z=\frac{A^2}{\alpha_0}(2x^2+1)^{-\frac{1}{4}}$$

or

$$y=A^2x(2x^2+1)^{-\frac{1}{4}}/\alpha_0. \quad (107)$$

If we plot y against x we have the required relationship between Δf_e and Δf_a. Curves of (105) and (107) are shown in Fig. 3 for different values of the minimum detectable signal α_0, $A=1$. The case $A=1$ is particularly important, as it enables us to compare the periodic and aperiodic metering schemes, where now $\Delta f_e=(\Delta f_m)_{\text{periodic}}$. See Section IV, Part I, and Figs. 3, 8 for a more detailed discussion.

We would like to express our sincere thanks for helpful comments and criticisms to Drs. E. R. Brill, W. H. Jordan, D. O. North, D. Taylor, and H. Wallman.

The Permeability of Different Rubbers to Gases and Its Relation to Diffusivity and Solubility*

G. J. VAN AMERONGEN
Rubber-Foundation, Delft, Holland
(Received June 6, 1946)

As the permeability of rubber-like substances to gases stands in relation to the solubility and rate of diffusion of the gases in those materials, these individual values should be known. The permeability of a membrane was measured manometrically and the diffusivity was derived from the time-lag of the permeation. The solubility of the gas was computed from the permeability and the diffusivity, in addition to which the solubility was also found by direct measurement. In this way eight different gases were tested with nine elastomers at different temperatures. It appeared that the permeability of a membrane to a given gas is not affected by the presence of a second gas. The differences in permeability of different elastomers to a given gas are caused mainly by differences in rate of diffusion and only in a very minor degree to differences in solubility. The differences in permeability of the same elastomer to different gases are caused not only by differences in rate of diffusion but also by differences in solubility. A linear relationship is found between the logarithm of the solubilities of different gases in natural rubber and their critical temperatures, so the higher the critical temperature of a gas, the better does it dissolve. The presence of polar groups in an elastomer reduces the solubility of non-polar gases and increases the solubility of polar gases in the elastomer. The various rubbers behave towards gases like organic liquids. The activation energy of the diffusion and the heat of solution were calculated from the temperature function of the diffusivity and the solubility. As the diameter of the molecule of the gas increases, the rate of diffusion decreases, while the activation energy of the diffusion becomes greater. The presence of polar groups and methyl groups in elastomers causes low rate of diffusion, which involves a great activation energy of diffusion. It is presumed that the activation energy of the diffusion is required to separate the rubber molecules for the displacement of the gas molecules. The attempt to elucidate the constant D_0 in the equation $D=D_0 \exp(-E/RT)$—which proves to be a function of the activation energy of the diffusion E—by reference to one of the formulas published in the literature failed. An empirical formula was drawn up, by which D_0 is directly related to the activation energy E.

1. INTRODUCTION

AS far back as 1831 Mitchell[1] discovered that a rubber membrane is permeable to gases and that the rate of permeation is different for different gases. With the advance of synthetic rubbers in recent years it furthermore became evident that there was a marked difference in the behavior of these various types of rubber in the presence of a given gas. While this divergence in the behavior of different gases and kinds of rubber is primarily of great technical importance (one has only to think of balloons and inner tubes), it also offers interesting aspects from the theoretical point of view.

All the experimental data up to date[2] go to show that, with rubber, the process of permeation is one of solution of the gas molecules on one side of the membrane followed by diffusion through the rubber to the other side, where the gas evaporates. Another fact which has emerged is that the quantity of dissolved gas in the rubber is proportional to the pressure, following Henry's law, while Fick's law applies to the actual process of diffusion. On this assumption, the quantity of permeating gas q is determined by[3]

$$q=DhA[(p_1-p_2)/d]t, \tag{1}$$

where

D = the diffusivity,
h = the solubility,
A = the area of the membrane,
d = the thickness of the membrane,
p_1 and p_2 = the pressures on both sides of the membrane,
and t = the time.

If unit values are taken for the difference in pressure p_1-p_2, the area A, the thickness d, and the time t, then

$$Q=Dh, \tag{2}$$

where Q stands for the permeability.

The units which will be used can be seen from this formula. According to *Bunsen*, the solu-

* Communication No. 55 of the Rubber-Foundation, Delft, Holland.

[1] J. V. Mitchell, J. Roy. Inst. **2**, 101, 307 (1831).

[2] Cf. H. A. Daynes, Trans. Inst. Rubber Ind. **3**, 428 (1927).

[3] S. von Wroblewski, Ann. d. physik. Chemie **8**, 29 (1879).

bility h is expressed by the number of cubic cm of gas of 0°C and 76 cm Hg which dissolves per cu. cm of rubber and 1 atmos. gas pressure. The diffusivity D is expressed in sq. cm per second and the permeability Q by the number of cu. cm of gas at 0°C and 76 cm Hg which penetrates 1 sq. cm of the surface of a piece of material 1 cm thick per second when the outside difference in pressure of the gas amounts to 1 atmosphere. The dimensions of Q are

$$[l^2t^{-1}p^{-1}] \quad \text{or} \quad [l^3m^{-1}t].$$

It is necessary to add that the above formulas apply only if the rate of solution and evaporation is very great in comparison to the rate of diffusion, in other words, if the concentration of the gas in the rubber interface remains constant and equal to the concentration of equilibrium (according to Henry's law), irrespective of the rate of diffusion. For instance, permeation is a more complicated phenomenon with metals, when the rate of solution or evaporation of the gas often governs the rate of permeation. This will be evident chiefly by the fact that the rate of permeation is no longer in inverse ratio to the thickness of the membrane. It has been experimentally proved, however, that formula (1) holds good for rubber, so with that material there is no reason to assume that the rate of solution has much influence upon the rate of permeation.

Our purpose now was to find out why a given kind of rubber possesses a distinctive permeability. To do this it will not suffice to make permeability measurements alone; it will be necessary to measure separately the factors, solubility and rate of diffusion of the gas in the elastomer, which, together, determine the permeability. It is best, moreover, to determine these two factors in dependence upon the temperature, since the heat of solution and the activation energy of the diffusion can then be calculated, by which means better insight into the whole process can be obtained.

2. METHOD FOR THE DETERMINATION OF THE PERMEABILITY

The apparatus with the aid of which the permeability and rate of diffusion were measured works on the same principles as that designed by De Boer and Fast[4] and by Barrer,[5] the characteristic feature of which is a permeability cell with a pressure gauge, as represented in Fig. 1.

A gas cell is made up of two parts of steel, A and B, containing gas chambers separated from each other by a rubber membrane. The gas is introduced into the A chamber, the volume of which is approximately 75 cu. cm, while the B chamber, which is connected to a pressure gauge, is exhausted to approximately 0.1 mm Hg pressure. To prevent the membrane from sagging, it is supported on fine gauze netting which, in turn, rests on a perforated iron table.

The flanges, when warmed, were given a thin coat of an asphaltic bitumen mixture so that, once they had been screwed down, the apparatus was completely gas-tight.

A McLeod pressure gauge, with a range up to 2.5 mm Hg pressure and recording with an accuracy of 0.005 mm Hg pressure, was used to measure the pressure in the vacuum chamber. By noting for some time the increase in pressure in the vacuum chamber due to the permeability of the rubber, the permeability Q can be found

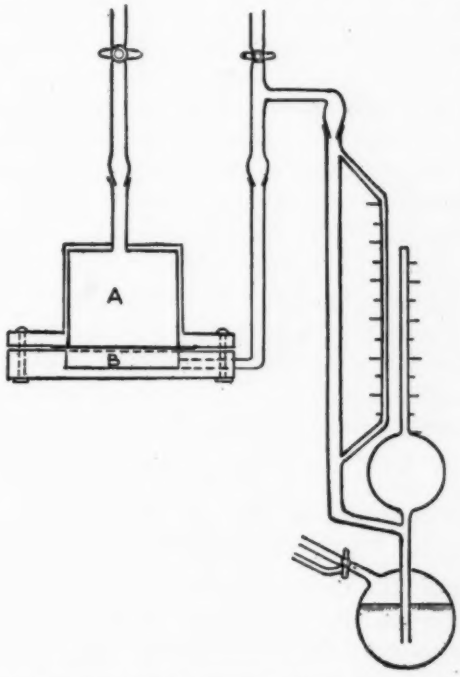


FIG. 1. Apparatus for the measurement of the permeation and diffusion of gases in rubber.

[4] J. H. De Boer and J. D. Fast, Rec. trav. chim. **57**, 317 (1938).
[5] R. M. Barrer, Trans. Faraday Soc. **35**, 628 (1939).

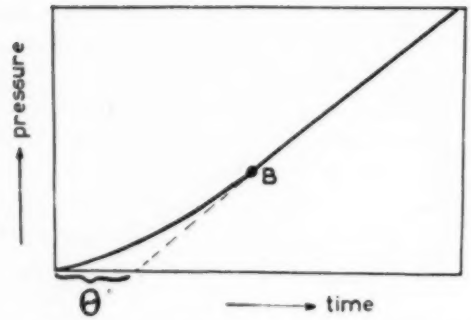


FIG. 2. Relation between increase in pressure and the time.

with the aid of the equation

$$Q=\frac{dp}{dt}\times\frac{V}{p}\times\frac{d}{A}\times\frac{273}{T}, \tag{3}$$

where

dp = increase in the pressure in B in time dt,
p = gas pressure in the gas chamber A,
V = volume of the apparatus on the vacuum side,
d = thickness of the membrane,
A = area of the membrane, and
T = average temperature of the vacuum chamber.

The surface area A of the membrane was always 30.2 sq. cm. The thickness of the membrane was determined with a Schopper thixometer to within an accuracy of 0.005 to 0.01 mm. The gas pressure p, which should be constant during a test, was usually 760 mm Hg. So that the measurements could be made at constant temperature, the gas cell was placed in a water thermostat. The membranes were made as described on a former occasion.[6]

This apparatus has much to recommend it for permeability determinations, which can often be carried out very quickly and with considerable recording accuracy; the consumption of gas per determination is negligible. The determination of Q is accurate to within 5 percent.

3. METHOD FOR THE DETERMINATION OF THE RATE OF DIFFUSION

Using the foregoing permeability apparatus, the rate of diffusion can be measured, in combination with the permeability determinations, by a method evolved by Daynes[7] and Barrer.[5] If a gas is brought into contact with one side of a gas-free rubber membrane, it will take some time before the gas has reached the other side, which can be observed experimentally if the pressure of the vacuum (pressure gauge) side is plotted against the time, as in Fig. 2.

It takes some time before there is constant increase in pressure, from which the permeability is computed. Daynes succeeded in finding an equation which correlates the time-lag extrapolated to the zero axis θ, after which permeation begins, the diffusivity D and the thickness d of the membrane, starting from the diffusion equation

$$\delta C/\delta t=D\delta^2C/\delta x^2. \tag{4}$$

The boundary conditions for the problem set are that:

1. for $x=0$ and all t $C=C_0$,
2. for $x=d$ and all t $C=0$,
3. for $t=0$ and all x $C=0$, and
4. A stationary condition prevails at $t=\infty$, so $\delta C/\delta t=0$.

From this an equation can be obtained which gives the concentration C of the gas in the membrane as a function of time t and place x;

$$C=C_0\left(1-\frac{x}{d}\right)-\frac{2C_0}{\pi}\sum_{n=1}^{n=\infty}\frac{1}{n}\sin\frac{n\pi x}{d}\exp\left[-\left(\frac{n\pi}{d}\right)^2Dt\right]. \tag{5}$$

The amount of material which at the place $x=d$ in time t passes through the area A is

$$q=AD\int_0^t-(dC/dx)_{x=d}dt. \tag{6}$$

After determination of dC/dx from (5), substitution in (6) and integration, we get

$$q=AD\left\{\frac{C_0t}{d}-\frac{C_0d}{6D}\sum_{n=1}^{n=\infty}1-\exp\left[-\left(\frac{n\pi}{d}\right)^2Dt\right]\right\}. \tag{7}$$

To the stationary stage applies

$$q=AD\frac{C_0t_1}{d}. \tag{8}$$

If $t=\theta+t_1$, then by combination of (7) and (8)

$$D=d^2/6\theta. \tag{9}$$

Thus Eq. (9) offers a simple means of calculating the diffusivity D from the thickness d of the

[6] G. J. van Amerongen, Rev. Gén. du Caoutchouc **21**, 50 (1944).
[7] H. A. Daynes, Proc. Roy. Soc. **A97**, 286 (1920).

membrane and the time-lag θ of the permeability determination.

It is very important to know at what point the stationary condition begins.

In this point (B of Fig. 2) the term

$$\sum_{n=1}^{n=\infty} \exp\left[-\left(\frac{n\pi}{d}\right)^2 Dt\right]$$

in Eq. (7) approaches zero,

$$\sum_{n=1}^{n=\infty} \exp\left[-\left(\frac{n\pi}{d}\right)^2 Dt\right] = \sum_{n=1}^{n=\infty} \exp\left[-\frac{n^2\pi^2 t}{6\theta}\right]$$

because $D=d^2/6\theta$,

$$\sum_{n=1}^{n=\infty} \exp\left[-\frac{n^2\pi^2 t}{6\theta}\right] = \exp\left[-\frac{\pi^2 t}{6\theta}\right] + \exp\left[-\frac{4\pi^2 t}{6\theta}\right] + \exp\left[-\frac{9\pi^2 t}{6\theta}\right] + \cdots. \quad (10)$$

If various values are put into (10) for t, it will be found that this term can practically be neglected after a time which is 2.5 to 3 times that of the time-lag, and indeed it is usually from this point that extrapolation has taken place to the zero axis.

In using Eq. (9) it has been assumed that the pressure in chamber B of Fig. 1, which at most reached 2 to 3 mm Hg, may be disregarded in comparison with the pressure in A, which usually amounted to 760 mm Hg. The determination of D is accurate to 5–10 percent.

The rate of diffusion through some materials is so slow that, as a result, the determinations sometimes take hours to carry out. For example, the time-lag of the permeation of N_2 through a membrane of Oppanol 0.20 mm thick is 40 minutes at 17°C, so that 100 to 120 minutes elapse before the stationary condition sets in.

It need hardly be said that the membrane has to be de-gassed after each test by evacuation for a period which is 3 to 4 times the time-lag.

An alternative to this method is to find the diffusivity as the quotient of the permeability and solubility, provided the solubility has been measured by some independent method, as described in the following section.

4. METHOD FOR THE DETERMINATION OF THE SOLUBILITY

It is possible to calculate indirectly the solubility of a gas in rubber from Eq. (2), which shows it to be the ratio of the permeability Q to the diffusivity D.

This calculation requires, of course, the independent determination of Q and D without the use of Eq. (2). A direct measurement of the solubility may be made with the apparatus shown in Fig. 3.

A flask A, filled with fine snippings of rubber, is connected to a gas volume meter BC sealed with mercury. After chambers A and B have been exhausted, cock K is closed and chamber B filled with the gas under test. The gas is raised to exactly one atmosphere pressure by means of a movable mercury tube C. Cock K is then opened and the gas streams from chamber B to A. After waiting until equilibrium is attained, constantly adjusting tube C in order to maintain pressure at 1 atmosphere, the gas volume is read off. This volume of gas, less the non-rubber volume of flask A, gives us the quantity of gas absorbed by the rubber. Determination can be made at various temperatures by suspending A up to cock K in a water thermostat.

With most gases the test is accurate to 5 percent. The solubility of CO_2 was determined to 1 to 2 percent accuracy, as this was the most suitable of the gases for the apparatus.

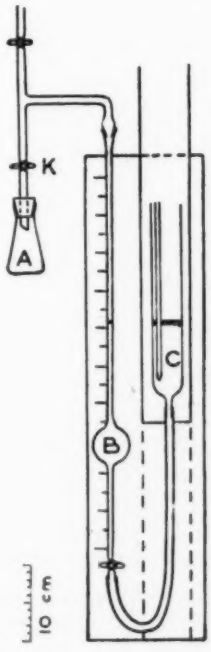


FIG. 3. Equipment for measuring the solubility of gases in rubber.

TABLE I. The permeability Q (in 10^{-8} cm² sec.$^{-1}$ atmos.$^{-1}$) and the diffusivity D (in 10^{-7} cm² sec.$^{-1}$) of gas in various elastomers.

	H_2			O_2			N_2			CO_2			CH_4		He
t in °C	Q	$D=\frac{d^2}{6\theta}$	$D=\frac{Q}{h}$	Q	$D=\frac{d^2}{6\theta}$	$D=\frac{Q}{h}$	Q	$D=\frac{d^2}{6\theta}$	$D=\frac{Q}{h}$	Q	$D=\frac{d^2}{6\theta}$	$D=\frac{Q}{h}$	Q	$D=\frac{Q}{h}$	Q
Natural rubber															
17°	28	79	76	12	12.5	12	4.1	8.0	8.3	72	6.7	7.1	12.5		16.5
25°	39	105	100	18	17.5	18	6.6	11.5	12.5	102	10.5	11	22	8.8	23
35°	58.5	140	140	28.5	27	28	11.0	20	21	145	17	18.5	36		33
43°	77	185	180	39	36	39	16	28	29	185	25	26	50	21	44
50°	97	220	230	49.5	49	49	22.5	37	40	220	32	34	64		
Buna S															
17°	22.5	80		9.0	9.6		3.0	7.2		71	6.8		10.5		13
25°	30.5	100	85	13	14	14	4.8	10	10	94	10	11	16		17.5
35°	44	135		20	20		7.8	14.5		130	15.5		26		25
43°	59.5	165	150	27.5	28	30	11.5	21	23	165	23	24	35		33
50°	74	200		34.5	34		14.5	28		195	29		43		42
Perbunan															
17°	7.6	31		2.0	2.4		0.50	1.45		15	1.0		1.3		5.8
25°	11.5	42	41	3.2	3.6	4.0	0.89	2.3	2.5	23	1.7	1.8	2.4		8.7
35°	17.5	64		5.3	6.3		1.65	4.1		37	3.1		4.7		12.5
43°	25.5	86	82	7.7	9.1	9.6	2.5	6.2	6.3	52	4.7	5.3	7.0		16.5
50°	31.5	110		10.5	13		3.7	8.6		66	7.0		10.1		21
Neoprene G															
17°	6.8	29		1.75	2.5		0.53	1.55		12.4	1.3		1.4		
25°	10.3	38	36	3.0	3.8	4.0	0.89	2.4	2.5	19.5	2.3	2.4	2.5		
35°	16	56		5.1	6.2		1.65	4.4		31	4.2		4.8		
43°	23	74	72	7.7	10.0	10.0	2.55	7.2	6.7	43.5	6.8	6.8	7.2		
50°	28.5	94		10.1	13		3.55	9.4		56.5	9.1		9.8		
Oppanol B 200															
17°	3.0	10		0.54	0.50		0.11	0.27		2.3	0.32		0.2		3.8
25°	4.9	14	14	0.90	0.78	0.84	0.22	0.43	0.42	3.8	0.54	0.55	0.55		5.6
35°	8.3	21		1.6	1.55		0.44	0.84		6.6	1.05		1.1		8.5
43°	12	31	30	2.6	2.4	2.5	0.78	1.5	1.4	10.0	1.8	1.8	1.9		12
50°	16.5	38		3.7	3.4		1.15	2.1		14	2.5		2.9		15.5
Butadiene rubber															
17°	23	75		10.1	11		3.4	8.1		80	7.6				
25°	32	96		14.5	15		4.9	11		105	10.5				
35°	46	125		21	22		7.8	16		140	16				
43°	60	160		28	30		11	22		175	22				
50°	77	180		36	37		14.5	29		200	28				
Methyl rubber															
17°	9.0	27		0.93	0.88		0.20	0.46		3.0	0.36		—		8.1
25°	13	39		1.6	1.4		0.36	0.79		5.7	0.63		0.60		11
35°	20	61		3.1	2.5		0.78	1.7		10.5	1.3		1.45		16
43°	29	80		5.0	4.1		1.3	2.8		17	2.2		2.7		22
50°	38	105		7.1	6.1		2.2	4.1		24	3.6		4.4		27
Mipolam MP															
17°	2.9	15		0.41	0.90					2.3	0.23				
25°	4.4	20		0.70	1.3		0.2			4.0	0.44				
35°	7.3	30		1.4	2.4		0.4			7.6	0.83				
43°	10.5	39		2.1	3.9		0.7			11.5	1.4				
50°	14	50		3.2	5.2					15	1.9				
Thiokol B															
17°	0.71	7.8		0.11						1.3	0.43				
25°	1.2	10.5		0.22						2.4	0.81				
35°	2.2	18		0.49						4.8	1.7				
43°	3.3	26		0.85						7.7	2.8				
50°	4.6	35		1.3						11.0	4.0				

5. EXPERIMENTAL TESTS

The rubbers used for the determinations were natural rubber, Buna S, Perbunan, methyl rubber (polydimethylbutadiene) and butadiene rubber (polybutadiene) vulcanized for 30 minutes at 142°C in the proportions of 100 parts of rubber, 1.5 parts of sulphur, 5 parts of zinc oxide, 2 parts of stearic acid, 1.8 parts of Vulkacite AZ, 1 part of phenyl-β-naphthylamine. Without any additions, Neoprene G was vulcanized for 15 minutes at 142°C, while Thiokol B was vulcanized with 10 parts of zinc oxide and 1 part of stearic acid for 25 minutes at 142°C. Thus any further mention of a particular kind of rubber *always* connotes a vulcanized one. This does not apply to Oppanol B 200 (polyisobutene) and Mipolam MP (vinylchloride-methylacrylate copolymer), which cannot be vulcanized and therefore required no admixtures, though, as a commercial product, Mipolam MP contains 27 percent of softener.

The principal tests were preceded by a few H_2 permeability determinations with natural rubber in dependence upon pressure and thickness in

TABLE II. The solubility h of gas in various elastomers.

	H_2		O_2		N_2		CO_2		CH_4	NH_3	SO_2
t in °C	Direct	$h=\frac{Q}{D}$	Direct	$h=\frac{Q}{D}$	Direct	$h=\frac{Q}{D}$	Direct	$h=\frac{Q}{D}$	Direct	Direct	Direct
Natural rubber											
17°	0.037	0.036	0.099	0.100	0.049	0.054	1.01	1.08	0.26		
25°	0.039	0.038	0.099	0.102	0.052	0.056	0.90	0.97	0.25	6.9	23.6
35°	0.041	0.040	0.100	0.104	0.053	0.058	0.78	0.83	0.25		
43°	0.042	0.042	0.100	0.105	0.055	0.059	0.70	0.74	0.24	4.1	11.6
50°	0.043	0.044	0.101	0.106	0.056	0.060	0.64	0.68	0.23		
Buna S											
25°	0.036	0.031	0.093	0.094	0.048	0.048	0.87	0.92		8.1	17.2
43	0.039	0.036	0.093	0.096	0.050	0.053	0.69	0.72		6.6	9.8
Perbunan											
25°	0.028	0.026	0.079	0.084	0.035	0.038	1.28	1.38		13.5	48.0
43°	0.031	0.029	0.080	0.084	0.040	0.041	0.98	1.08		8.0	23.5
Neoprene G											
25°	0.029	0.026	0.075	0.075	0.036	0.036	0.83	0.83		8.8	18.1
43°	0.032	0.030	0.076	0.078	0.038	0.037	0.64	0.66		6.7	10.5
Oppanol B 200											
25°	0.035	0.034	0.107	0.110	0.052	0.049	0.69	0.69		1.25	3.6
43°	0.040	0.041	0.102	0.108	0.054	0.052	0.57	0.57		1.05	2.4
Butadiene rubber											
25°		0.034		0.094		0.045		0.97			
43°		0.039		0.097		0.049		0.82			
Methyl rubber											
25°		0.033		0.114		0.045		0.87			
43°		0.035		0.117		0.050		0.72			
Mipolam MP											
25°		0.022		0.052				1.00			
43°		0.027		0.059				0.80			
Thiokol B											
25°		0.011						0.30			
43°		0.013						0.28			

order to ascertain whether the quantity of permeating gas is proportional to the pressure and in inverse ratio to the thickness, as stated in the literature.[8] This proved to be correct, in measurements made between 160 and 1700 mm Hg pressure with membranes of 0.20 to 1.01 mm thickness.

TABLE III. The specific gravity and the cubic expansion coefficient λ of various elastomers.

	Sp. gr. at 15°C	λ between 15° and 40°C
Natural rubber (vulcanized)	0.9689	7.0×10^{-4}
Buna S (vulcanized)	0.9974	6.7
Perbunan (vulcanized)	1.0316	6.2
Neoprene G (vulcanized)	1.2470	6.2
Oppanol B 200	0.9306	6.2

[8] J. D. Edwards and S. F. Pickering, Bur. Stand. Sci. Paper 387 (1920).

The results of the main work are recorded in Tables I and II.

The thickness of the membranes, which varied from approximately 0.2 to 1.1 mm, was always so chosen as to produce a measurable time-lag. Of the gases used, H_2, O_2, N_2, SO_2, and NH_3 were obtained from cylinders, the gas CO_2 being taken from a Kipp apparatus. The methane was prepared by allowing water to drip into the Grignard compound of CH_3I with Mg.

To determine the solubility by the direct method it is necessary to know the specific gravity and the cubic expansion coefficient of the rubbers used. The specific gravity was therefore determined at a number of different temperatures by the buoyancy method, *viz.*, by weighing on an analytical balance, the rubber, suspended

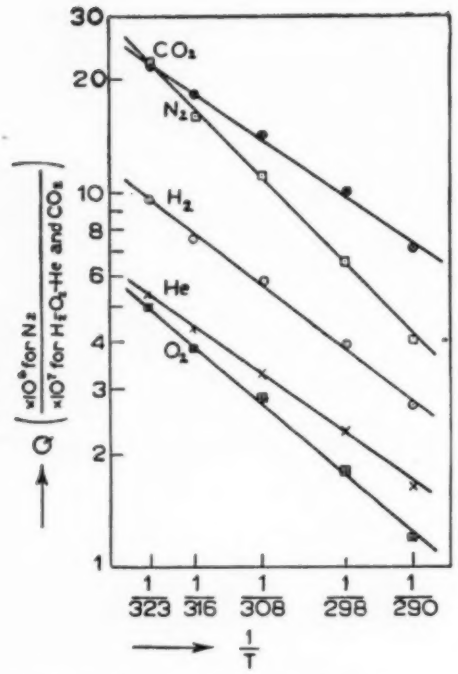


FIG. 4. Relation between the permeability in natural rubber and temperature.

on a hook, being dipped under water. The results are shown in Table III.

The specific gravity is accurate to 0.001 and λ to 0.2×10^{-4}. Within the measured range the coefficient of expansion is not a function of the temperature.

The computed solubilities contain the error of the permeability and diffusion determinations. To reduce this error to a minimum, the solubility was not calculated direct from the quotient of the permeability and diffusivity, but the latter quantities were first plotted logarithmically against $1/T$, as in Figs. 4 and 5. From this the solubility was calculated graphically. This was a precautionary measure against two accidental errors producing a considerable error in the solubility. As these solubilities are accurate to 10 percent, the results obtained by the direct method are the more reliable.

The permeabilities for H_2 and CO_2 at 25°C agree within the experimental error with the permeabilities determined by other means in the presence of air.[6] It follows from this that it is immaterial whether there is a vacuum or some alien gas on the other side of the membrane, as was the case in the former experimental tests. Thus permeability to a given gas is not affected by the presence of an alien gas. There is, likewise, agreement between solubilities and diffusivities, which were determined in two ways in this work.

For practical purposes the permeabilities at 25°C of the various elastomers are calculated with reference to natural rubber in Table IV.

According to our results, the differences in permeability to a given gas are caused mainly by differences in rate of diffusion and only in a very minor degree by differences in solubility.

In Table V the permeabilities at 25°C have been referred to H_2. The differences in the permeability of a given elastomer to various gases are caused not only by great differences in rate of diffusion, but also to a considerable extent by differences in solubility. For example, the permeability to CO_2 is as a rule far greater than to H_2 owing to the far greater solubility of CO_2, which amply compensates for its far slower rate of diffusion.

6. CALCULATION OF THE ACTIVATION ENERGY OF THE PERMEATION, THE ACTIVATION ENERGY OF THE DIFFUSION AND THE HEAT OF SOLUTION

The thermal dependence of the permeability and the diffusivity takes an exponential course, as it appears that there is a linear relation between $\log Q$ or $\log D$ and $1/T$. Some of the results plotted in this way are shown in Figs. 4 to 6, from which the applicability of the fol-

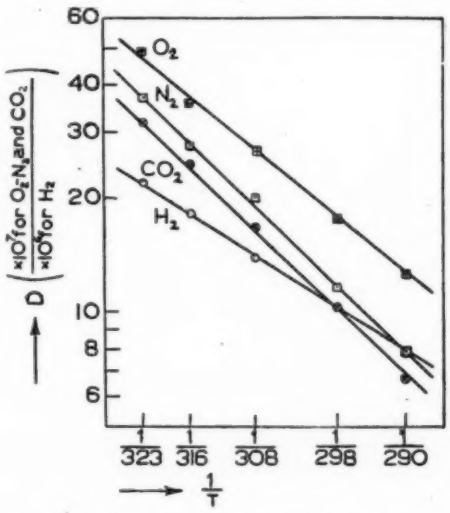


FIG. 5. Relation between the diffusivity in natural rubber and temperature.

lowing formulas may be concluded:

$$Q = Q_0 e^{-W/RT} \tag{11}$$

and

$$D = D_0 e^{-E/RT} \tag{12}$$

where Q_0 and D_0 = constants, W = activation energy of the permeation, and E = activation energy of the diffusion.

The slope of the lines in Figs. 4 to 6 gives a measure of the energy. The constants and the energies can be calculated with the aid of formulas (13) and (14), which follow from formula (12).

$$E = \frac{4.573 \log D_2/D_1}{1/T_1 - 1/T_2}, \tag{13}$$

$$\log D_0 = \log D_1 + \frac{E}{4.573 T_1}. \tag{14}$$

Dividing (11) by (12), we get

$$\frac{Q}{D} = \frac{Q_0}{D_0} e^{(-W+E)/RT}, \tag{15}$$

or

$$h = h_0 e^{(-W+E)/RT}. \tag{16}$$

And it is indeed a fact that the solubilities found which, in the case of natural rubber, were determined in complete thermal dependence, prove to take an exponential course with the temperature.

TABLE IV. The permeability of elastomers at 25°C with reference to natural rubber.

	H_2	O_2	N_2	CO_2	CH_4	He
Natural rubber	100	100	100	100	100	100
Butadiene rubber	82	80	74	103	—	—
Buna S	78	72	73	92	73	76
Methyl rubber	33	9	5.5	5.6	2.7	48
Perbunan	29	18	14	23	11	38
Neoprene G	26	17	14	19	11	—
Oppanol B 200	12	5	3.3	3.7	2.5	24
Mipolam MP	11	4	3	3.9	—	—
Thiokol B	3	1.2	—	2.4	—	—

TABLE V. The permeability of elastomers at 25°C with reference to H_2.

	H_2	O_2	N_2	CO_2	CH_4	He
Natural rubber	100	46	17	260	56	59
Butadiene rubber	100	45	15	330	—	—
Buna S	100	43	16	310	52	57
Methyl rubber	100	12	2.8	44	4.6	85
Perbunan	100	28	7.7	200	21	76
Neoprene G	100	29	8.6	190	24	—
Oppanol B 200	100	18	4.5	77	11	114
Mipolam MP	100	16	4.5	91	—	—
Thiokol B	100	18	—	200	—	—

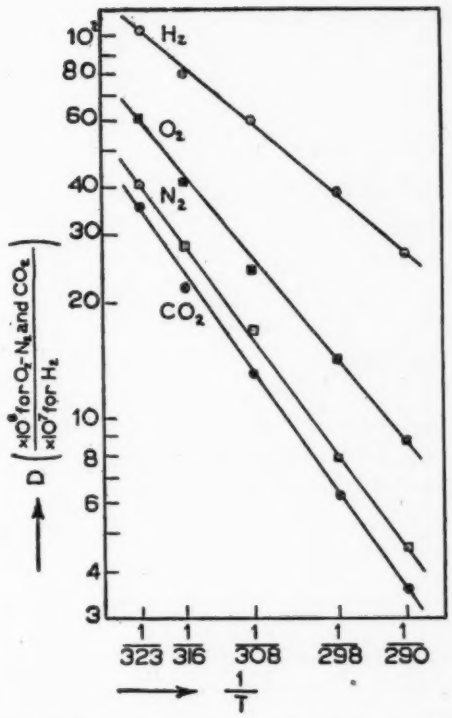


FIG. 6. Relation between the diffusivity in methyl rubber and temperature.

The following formulas hold for the calculation of the heat of solution:

$$\frac{d \ln C_B/C_A}{dT} = \frac{H}{RT^2}, \tag{17}$$

where H = heat of solution, C_B = concentration of the gas in the rubber, and C_A = concentration of the gas in the gas phase,

$$\frac{C_B}{C_A} = K_0 e^{-H/RT}. \tag{18}$$

Therefore, to calculate this heat of solution, the proportion of the concentration of dissolved gas to the concentration of the gas in the gas phase must be known, i.e., the number of cu. cm of gas of the prevailing temperature and pressure which is dissolved per cu. cm of solvent. It is quite simple to calculate this by multiplying our solubility figures, which are expressed in cu. cm of gas of 0°C and 76 cm by $T/273$.

The calculation then becomes as follows:

$$\ln h\frac{T}{273}=-\frac{H}{RT}+\ln K_0, \tag{19}$$

$$\frac{d\ln h}{d1/T}=-\frac{H}{R}+T=-\frac{H'}{R}, \tag{20}$$

$$H=RT+H'. \tag{21}$$

Thus, to find the required heat of solution, one can multiply the solubility h by $T/273$ and plot the logarithm of this amount against $1/T$. Alternatively, the logarithm of the solubility h can simply be plotted against $1/T$ and RT cal. added to the "heat of solution" H' thus obtained.

RT cal. has likewise to be added to $W-E$ to calculate the heat of solution from the temperature function of the permeability and diffusivity.

The heat of solution obtained in this way represents the supply of heat necessary to dissolve 1 mole of gas in rubber saturated with gas.

Table VI shows the activation energy of the permeation W for the various processes of permeation with the conjoined constant Q_0. Measurements made by Barrer[8] and Sager[9] show satisfactory agreement with these figures. It will be evident that both quantities have little intrinsic value. But the energy W was used for the calculation of the heat of solution, and it also permits of the calculation of the permeability Q for any desired temperature. Caution is, however, necessary when calculating the permeability in this way at temperatures well outside those of our measured range, as the energy W may be a function of the temperature; indeed, with CH_4, this already proved to be the case within that range, for which reason the activation energy W of CH_4 was not calculated. At lower temperatures the increase in permeability with the temperature was greater than at higher temperatures. With CO_2 as well, a similar tendency to depart from the purely exponential course was observed within the measured range, but with this gas the discrepancies were usually too small to affect the calculation of the activation energy W.

Table VI also gives the activation energy of the diffusion and the heat of solution of gases in the elastomers under test. In the case of natural rubber, moreover, the heat of solution is calculated from the direct determination of the solubility. The activation energy of the diffusion followed from that heat of solution together with the activation energy of the permeation. The figures resulting from these two forms of calculation agree.

The determination of the activation energies of the permeation and of the diffusion is accurate to 200 cal. The reason why this is so much more accurate than measurement of the permeability and diffusivity alone is that the calculation of these energies does not necessitate the use of various particulars which themselves contain errors, like the thickness and area of the membrane and volume of the apparatus, since all

TABLE VI. The activation energies W and E with the conjoined constants Q_0 and D_0 and the heat of solution H in cal./mole.

		Q_0	W	D_0	E	H
Natural rubber I (from Q and D determination)	H_2	0.046	6900	0.23	5900	1600
	O_2	0.080	6600	0.57	7500	800
	N_2	0.47	9300	2.9	8700	1200
	CO_2	0.036	6200	3.6	8900	−2100
	He	0.014	6500	—	—	—
Natural rubber II (from Q and h determination)	H_2	—	—	0.31	6100	1400
	O_2	—	—	0.66	7600	700
	N_2	—	—	2.8	8600	1300
	CO_2	—	—	3.3	8800	−2000
	CH_4	—	—	—	—	−1300
Buna S	H_2	0.031	6800	0.056	5100	2300
	O_2	0.030	7300	0.23	7100	800
	N_2	0.12	8700	0.55	7800	1500
	CO_2	0.014	5700	0.90	8100	−1800
	He	0.013	6600	—	—	—
Perbunan	H_2	0.099	8100	0.96	7300	1400
	O_2	0.25	9400	2.6	9300	700
	N_2	1.1	11000	7.2	10200	1400
	CO_2	0.29	8300	12	10700	−1800
	He	0.012	7000	—	—	—
Neoprene G	H_2	0.090	8100	0.28	6600	2100
	O_2	0.54	9900	3.1	9400	1100
	N_2	0.56	10600	9.3	10300	900
	CO_2	0.33	8500	20	10800	−1700
Oppanol B 200	H_2	0.47	9500	0.55	7600	2500
	O_2	0.66	10700	7.3	10800	500
	N_2	5.5	12800	18	11700	1700
	CO_2	0.98	10100	22	11700	−1000
	He	0.035	7900	—	—	—
Butadiene rubber	H_2	0.022	6600	0.053	5100	2100
	O_2	0.023	7100	0.15	6800	900
	N_2	0.053	8200	0.22	7200	1600
	CO_2	0.0069	5200	0.24	7300	−1500
Methyl rubber	H_2	0.10	8000	1.3	7500	1100
	O_2	3.3	11300	20	11100	800
	N_2	22	13300	105	12400	1500
	CO_2	9.7	11200	160	12800	−1000
	He	0.0078	6600	—	—	—
Mipolam MP	H_2	0.15	8900	0.20	6800	2700
	O_2	3.2	11800	4.3	10200	2200
	CO_2	0.68	9800	27	12000	−1600
Thiokol B	H_2	0.74	10600	2.0	8500	2700
	O_2	18	13500	—	—	—
	CO_2	6.9	11500	67	12100	0

[9] T. P. Sager, J. Research Nat. Bur. Stand. **25**, 309 (1940).

that is needed to calculate W is to plot the logarithm of the pressure increase per unit of time (after conversion to 0° and 76 cm) against $1/T$. It likewise suffices for the calculation of E to plot the logarithm of the time-lag against $1/T$. In unfavorable cases, i.e., when the errors in W and E reinforce each other, the determination of the heat of solution is faulty to 400 cal. The accuracy of Q_0 and D_0 is entirely governed by that of W and E. An error of 200 cal. in W or E makes Q_0 or D_0 faulty by approximately 30 percent.

7. DISCUSSION OF THE SOLUBILITY

With the data we have collected, we shall now consider more closely the two factors which govern permeability, *viz.*, solubility and rate of diffusion, in an endeavor to find the reason for the magnitude of the permeability of rubber to gases. It will become apparent that the behavior of gases in rubber is entirely analogous to that of gases in organic liquids. So, rubber may be regarded as an organic liquid of very high molecular weight.

One of the striking facts which emerges from the solubility measurements is that the solubility of H_2 and N_2 in rubber-like substances increases as the temperature increases, from which it follows that heat must be supplied for the process of dissolving. This is contrary to Barrer's[5] experience upon determining the solubility of H_2 and N_2 in some elastomers, when the solubility decreased as the temperature rose. Apart from the fact that we established the dependence of the solubility upon the temperature in two unallied ways, our finding is the more probable because the solubility of these gases in organic liquids almost invariably increases together with the temperature.[10] More soluble gases, like CO_2 for example, always exhibit a decrease in solubility, on the contrary, in organic liquids as the temperature rises.

The explanation we suggest for this behavior of gases is the following:[11]

The heat of solution H consists of a thermal value Q, depending upon the chemical or physical action between gas and liquid molecules, and of a work A brought about by the passing of 1 mole of gas into the solution, hence

$$H=A+Q, \tag{22}$$

whereby the heat absorbed during solution and the work then exerted by the gas become positive factors in the calculation. Furthermore

$$A=RT\ln P/p, \tag{23}$$

A being regarded as the work produced upon the conversion of 1 mole of gas from outside pressure P to the osmotic gas pressure p in the liquid. It should be noted that p/P is the solubility of the gas expressed as cu. cm gas of the prevailing temperature dissolved per cu. cm rubber.

If the concentration of the gas in the liquid is smaller than outside, A is positive. If Q is neglected, then $A=H$, with the result that the heat of solution H, which has to be supplied, is, likewise, positive, and solubility rises as the temperature rises.

If, on the other hand, the concentration of the gas in the liquid is greater than outside, A will be negative. Neglecting Q, it follows from this that the heat of solution H is negative and that the solubility will decrease as the temperature rises.

Reference to Table VI will show that the heat of solution of a gas in elastomers is less positive or more negative in proportion to greater solubility. It has been found that the heat of solution of N_2 and H_2 in the various rubbers agrees fairly closely with the work of expansion as calculated from (23). The thermal value Q cannot be neglected in the case of the other gases. Taking CO_2 as an instance, its solubility is of a kind that the concentration of the gas inside and outside the rubber is about equal; hence the work A is very small, yet an appreciable negative value is found for the heat of solution.

The degree of solubility is governed by the nature of the gas and of the elastomer.

a. The Nature of the Gas

The fact is known that a gas with a high critical temperature or boiling point dissolves more readily in an organic liquid than a gas of

[10] Cf J. Horiuti, Sci. Pap., Inst. Phys. Chem. Research, Tokyo **17**, 125, 256 (1931).
[11] G. Tammann, Zeits. f. anorg. Chemie **158**, 17 (1926); **194**, 159 (1930).

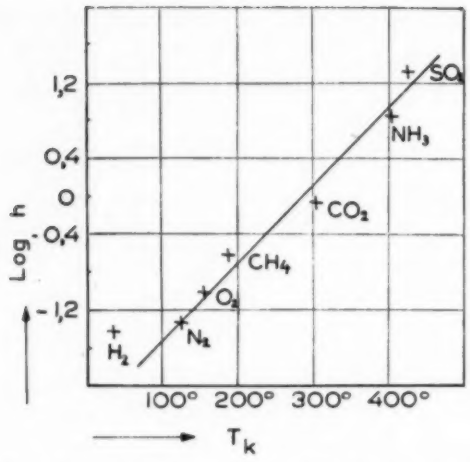


FIG. 7. Relationship between the critical temperature and solubility of gases in natural rubber.

lower critical temperature or boiling point.[12] The position of the critical temperature is a measure of the tendency of the gas to condense, with which the solubility stands in relation.

With rubbers, too, the solubility of the gas increases with the critical temperature.[13] As may be seen from Fig. 7, a relation exists between the solubility h and the critical temperature of the gas T_k:

$$\log h = A + BT_k, \qquad (24)$$

where A and B are constants.

H_2 is somewhat more soluble in rubber than would seem to follow from (24).

Körösy[14] found the same relation for the solubility of gas in organic liquids, with H_2 showing a similar deviation.

b. The Nature of the Elastomer

The presence of polar groups reduces the solubility of a non-polar gas in an organic liquid,[15] for two electrically polar molecules will have greater mutual attraction than two which are not polar. In a solution of non-polar molecules in a polar solvent this greater mutual attraction will more or less expel the non-polar molecules. Contrarily, polar gas molecules will dissolve more readily in a liquid with polar molecules than in a liquid without polar molecules.

All this is equally applicable to rubber-like substances. Perbunan with strong polar groups (−CN groups) has a reduced solvent power with respect to non-polar gases—H_2, N_2 and O_2—and an increased solvent power upon the polar gases SO_2 and NH_3, while on the other hand the saturated hydrocarbon Oppanol has very low solvent power over the polar gases. Perbunan also has an increased solvent power with respect to CO_2, despite the fact that this gas has no dipole moment as have SO_2 and NH_3. Nevertheless, CO_2 is not non-polar, for it contains two polar groups, which neutralize each other exteriorly,[16] one result of which is that, on the whole, CO_2 dissolves more satisfactorily in a substance with polar groups than in one without them, so in this case in Perbunan more than in natural rubber. The same thing is not found, however, with Neoprene (containing −Cl groups).

8. DISCUSSION OF THE ACTIVATION ENERGY OF DIFFUSION

The value of the diffusivity is governed by the value of the constant D_0 and the activation energy of the diffusion E. The activation energy E has to be supplied to separate the rubber chains—which are linked together by Van der Waals forces[17]—a sufficient distance to permit a gas molecule to be displaced or to make for itself a new place or hole. This energy is supplied by the kinetic energy of the gas molecule and particularly by the kinetic energy of the rubber molecules.

The degree of activation energy is governed by the nature of the gas and of the elastomer.

a. The Nature of the Gas

It might be expected that the larger the diameter of the gas molecule, the greater will the activation energy be, since larger openings between the rubber molecules will be needed for the diffusion of the larger gas molecules; consequently, these rubber molecules will need to be separated over a greater length, for which more energy is required.

[12] J. H. Hildebrand, *Solubility of Non-Electrolytes* (Reinhold Publishing Corporation, New York, 1936), second edition, p. 32.
[13] Cf. C. S. Venable, and T. Fuwa, Ind. Eng. Chem. **14**, 139 (1922).
[14] F. Körösy, Trans. Faraday Soc. **33**, 416 (1937).
[15] J. H. Hildebrand, reference 12, p. 133.
[16] G. Briegleb, *Zwischenmolekulare Kräfte und Molekülstruktur* (F. Enke Verlag, Stuttgart, 1937), pp. 50 and 61.
[17] R. H. Ewell, J. App. Phys. **9**, 252 (1938).

Figure 8 shows the activation energy in relation to the diameter d_T of the gas molecules used,[18] which was determined from the internal friction. It appears that the activation energy of the diffusion approximately increases in proportion to the size of the gas molecules.

b. The Nature of the Elastomers

It is to be expected that the stronger the secondary valencies by which the rubber molecules are interlinked, the greater will be the activation energy of the diffusion E. These secondary valencies also govern the extent of the heat of evaporation and the cohesive energy of the molecules.

The introduction of methyl groups into rubber causes the cohesive energy to increase, as may be seen in Table VII, where each successive elastomer has one methyl group per structural unit more than the previous elastomer. It is evident that the activation energy and the cohesive energy are interrelated. This cohesive energy applies per unit of structure and is calculated by Dunkel's method.[19] We may conclude that approximately $1\frac{1}{2}$ to $2\frac{1}{2}$ units of structure have to be separated to allow the diffusion of one gas molecule.

Oppanol is another instance of the considerable importance of the methyl groups in the molecule, for in it, as in methyl rubber, the activation energies for the diffusion were found to be very high. It is not possible in this case to calculate the cohesive energy following Dunkel, because the increment for this energy of a quaternary C atom is not known, but, as compared with natural rubber, Oppanol may be expected to possess considerable cohesive energy per linear unit of the chain molecule, as the Oppanol molecules contain a great many methyl groups.

TABLE VII. Comparison between the cohesive energy and the activation energy E (in cal./mole).

	Cohesive energy	E_{H_2}	E_{O_2}	E_{N_2}	E_{CO_2}
Butadiene rubber	3960	5100	6800	7200	7300
Natural rubber	4370	5900	7500	8700	8900
Methyl rubber	4780	7500	11100	12400	12800

[18] H. A. Stuart, *Molekülstruktur* (Verlagsbuchhandlung Julius Springer, Berlin, 1934), p. 36.
[19] M. Dunkel, Zeits. f. physik. Chemie **A138**, 42 (1928).

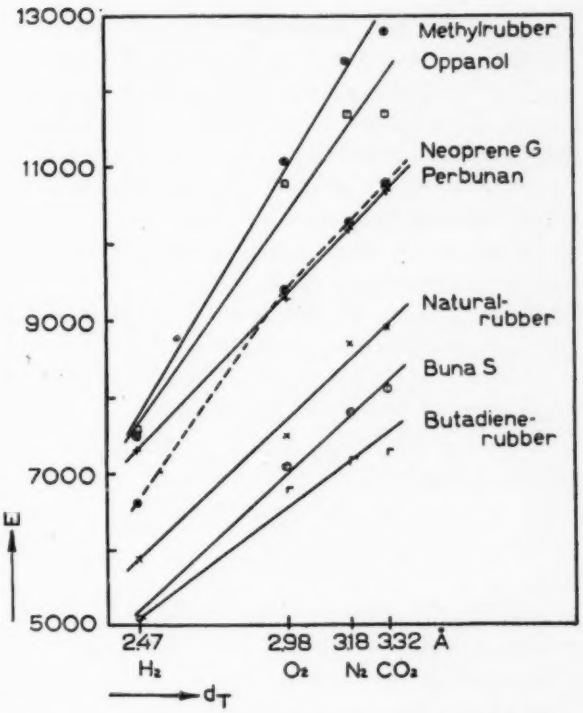


FIG. 8. The activation energies of the diffusion in relation to the diameter of the gas molecules.

The introduction of polar groups greatly increases the cohesive energy of molecules. Accordingly, a high activation energy of the diffusion is found in Neoprene and Perbunan as compared with butadiene rubber.

To sum up, it may be said that for the investigated materials a considerable activation energy of the diffusion, and thereby a low rate of diffusion and low permeability, may be brought about by the presence in the elastomer of (*a*) polar groups, e.g., Perbunan, Neoprene, Mipolam, and Thiokol; (*b*) methyl groups, e.g., methyl rubber and Oppanol.

9. DISCUSSION OF THE CONSTANT D_0

Several attempts have been made in the literature[20] to relate the temperature function of the diffusivity to other quantities. Most of the formulas describing this relation are of the following type:

$$D = \tfrac{1}{6}\nu d^2 e^{-E/RT}, \qquad (25)$$

where ν is a frequency and d a length.

[20] R. S. Bradley, J. Chem. Soc. 1910 (1934); Trans. Faraday Soc. **33**, 1185 (1937). R. M. Barrer, Proc. Phys. Soc. London **52**, 58 (1940).

The factor $\frac{1}{6}\nu d^2$ then corresponds to the constant D_0 as determined by us. If the frequency ν is conceived as the number of vibrations executed by the gas molecule in the rubber and the length d as the free length of path of this gas molecule in the rubber, it would follow that D_0 must be roughly constant in the diffusion of a given gas in various rubbers. Actually there is great disparity between the D_0's. It was found, for example, that for N_2 $D_0=0.22$ to 105, depending very much upon the activation energy E.

This makes it plain that a formula of this type does not hold good for rubber.

There is another formula, suggested by Barrer,[21] which might meet the case, *viz.*,

$$D=\tfrac{1}{6}\nu d^2\frac{(E/RT)^{f-1}}{(f-1)!}e^{-E/RT}, \qquad (26)$$

f = number of degrees of freedom.

This formula only applies if $(E/RT)\ll f-1$.

If this formula is applied to our results, then E has been wrongly calculated, since the fact that D_0 may depend upon the temperature has not been taken into account, for

$$\frac{d\ln D}{d1/T}=-\frac{E}{R}+(f-1)T=\frac{-E(\text{determined})}{R}. \qquad (27)$$

Therefore, $(f-1)RT$ has to be added to the E already determined and this new E must be filled in in (26). Barrer worked out the results of his diffusion measurements of gas in rubber with this formula and came to the conclusion that, to obtain a proper order of magnitude for D_0, the number of degrees of freedom must amount to 12 to 14.

We worked out some of our results with this formula, assuming that $\nu=5\times10^{12}$ and that $d=8$A. It makes no fundamental difference if some other amount is taken for these quantities.

TABLE VIII. Calculation of the number of degrees of freedom f in accordance with formula (26).

	D_0	E	Calculated	
H_2—Buna S	0.056	5100	$f=4$	$D_0=0.051$
H_2—Natural rubber	0.23	5900	6	0.21
H_2—Perbunan	0.96	7300	8	0.93
N_2—Oppanol	18	11700	10	22
N_2—Methyl rubber	105	12400	13	96

[21] R. M. Barrer, Trans. Faraday Soc. **35**, 644 (1939).

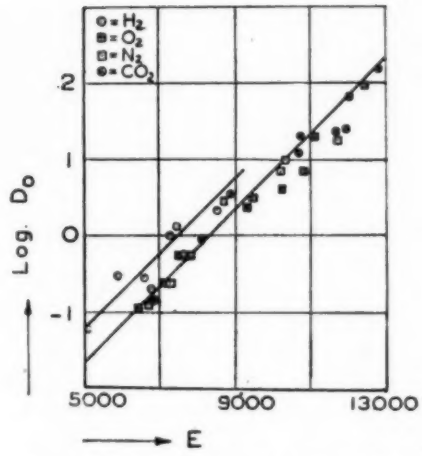


FIG. 9. Relation between log D_0 and E.

It appears from Table VIII that the number of degrees of freedom may vary from 4 to 13, according to the activation energy, which does not clarify the problem at all. It is difficult to avoid the impression that f here is little more than a factor by which D_0 may be brought into the right order of magnitude.

So, as this course does not provide satisfactory information on the magnitude of D_0, we tried to draw up an experimental formula based on the great dependence of D_0 upon E. For this purpose all the D_0 values found were logarithmically plotted against E (see Fig. 9), with the result that a practically linear relation is found to exist between log D_0 and E. Barrer[22] also found this relation.

The points in Fig. 9, fall fairly clearly into two groups. The first group, with O_2, N_2, and CO_2, follows by approximation a line which conforms to the equation

$$\log D_0=E/2000-4.2$$

or

$$D_0=6.3\times10^{-5}e^{E/870}. \qquad (28)$$

The second group with H_2 conforms by approximation to

$$\log D_0=E/2000-3.75$$

or

$$D_0=1.8\times10^{-4}e^{E/870}. \qquad (29)$$

It also follows from these formulas that H_2 diffuses on an average about 2 to 3 times more

[22] R. M. Barrer, Trans. Faraday Soc. **38**, 322 (1942).

quickly than O_2, N_2, or CO_2, apparently on account of the greater average velocity of H_2 molecules. The kind of rubber also seems to exert a specific influence, as gas diffuses in natural rubber approximately twice as quickly and, in Oppanol and Mipolam, about twice as slowly as is to be inferred from the formulas. The combination of the formulas (28) and (29) with $D=D_0e^{-E/RT}$ leads for N_2, O_2, and CO_2 to

$$D=6.3\times10^{-5}\exp\left[-E\left(\frac{1}{RT}-\frac{1}{870}\right)\right], \quad (30)$$

and for H_2 to

$$D=1.8\times10^{-4}\exp\left[-E\left(\frac{1}{RT}-\frac{1}{870}\right)\right]. \quad (31)$$

Hence these formulas apply by approximation to all species of rubber and the first one even to three gases. According to (30) and (31), the diffusivity D depends only, with a given gas, upon E and the temperature; thus any random rubber could be roughly classified merely by measuring the rate of diffusion at a certain temperature. Indeed, with only one exception, the higher activation energies E were always found to go along with the lower diffusivity D.

There would be some justification for concluding from this reasoning that the considerable differences in the e-power are responsible for the great differences in the D_0's found and it may therefore be asked whether the differences in D_0 have any real significance.

The most serious problem with which one is faced when drawing up a formula intended to meet all theoretical and experimental demands is that of accounting for the heat motion of the rubber molecules, and the failure of diffusion formulas which do satisfy other diffusion processes is probably caused by this inadequate allowance for the heat motion. This view accords with the results of an investigation carried out by Kuhn[23] into the rate of diffusion of organic dyes in swollen and de-swollen rubber, where, likewise, the heat motion of the rubber molecules proved to be of decisive importance.

ACKNOWLEDGMENT

In conclusion, I wish to express my thanks to the Rubber-Foundation for permission to publish the results of the investigation; to Professor Dr. A. van Rossem for his interest and advice; and to Professor Dr. C. Zwikker for his collaboration in the preparation of the thesis which was published in Delft in 1943 and of which this paper reproduces a part.

[23] W. Kuhn, H. Suhr, and K. Rijffel, Helv. phys. acta **14**, 497 (1941).

Here and There

New Appointments

Donald Kirkham has accepted a position as associate research professor in the Department of Agronomy and Physics at Iowa State College, where he will be in charge of the soil physics work.

William H. Myers has been appointed chief engineer of the receiver division, Farnsworth Television and Radio Corporation.

Carl J. Christensen of the Bell Telephone Laboratories technical staff has been appointed dean of the School of Mineral Industries at the University of Utah.

R. O. Curry is now audio and acoustical engineer for the Farnsworth Television and Radio Corporation. He was previously engaged in audio research for the Capehart Division of the company.

Willard R. Beye, formerly senior engineer at Transcontinental and Western Airlines, Kansas City, Kansas, has joined the staff of the Midwest Research Institute. He will work in the design and development section. Y. W. Luke, computer and mathematician, has also become associated with the Institute.

Howard P. Wile is the first appointee to the new post of administrator of research for contract projects at the Polytechnic Institute of Brooklyn.

Frank H. Spedding has been appointed director of the newly-established Atomic Research Institute at The Iowa State College, Ames, Iowa.

New Section for Edison Laboratory

The Edison Laboratory, central research organization of Thomas A. Edison, Incorporated, West Orange, New Jersey, has announced the formation of a new laboratory section for fundamental research in the branches of science which are the bases of the company's business in dictating machines, batteries, electrical instruments, and other electrical devices. The technical staff of the new group includes Dr. Robert S. Weisz, inorganic chemist, Dr. Richard W. Iskraut, physicist, S. S. Jaffe, chemist, and Dr. Robert L. Tichenor, physical chemist.

Radio Technical Planning Officers

The following men have been elected to serve as officers of the Radio Technical Planning Board for the year beginning October 1: chairman, HARADEN PRATT of Mackay Radio and Telegraph Company; vice chairman, J. L. MIDDLEBROOKS of the National Association of Broadcasters, Inc.; secretary, GEORGE W. BAILEY, executive secretary of the Institute of Radio Engineers, Inc.; treasurer, WILL BALTIN of the Television Broadcasters Association.

Bausch and Lomb Scholarships Awarded

Competition among fifteen finalists for five Bausch and Lomb Science Scholarships at the University of Rochester, valued at $1500 each, was so keen that the University has awarded a total of eleven scholarships, with an aggregate maximum value of more than $16,000. The finalists were chosen from nearly 450 contestants among the highest ranking high school science students in the United States, and were brought to Rochester, New York, for final tests and interviews. This was the third annual nationwide competition for these scholarships. Eight boys and three girls were the winners.

Appeal for Better Prosthetic Devices

Dr. Paul E. Klopsteg, chairman of the Committee on Prosthetic Devices, has issued, through the facilities of the National Inventors Council, a challenge to American inventors and mechanical engineers to develop better artificial limbs. Stressing the urgent needs of 17,000 war amputees and an estimated annual minimum of 25,000 amputees, Dr. Klopsteg urged inventors and engineers to submit ideas and suggestions for better devices to John C. Green, Chief Engineer, National Inventors Council, Department of Commerce, Washington 25, D. C. He listed the following pressing problems as still requiring solution: improved joint mechanisms, lighter materials, new fabrication methods, fundamental study of the mechanics of human motion, improved fitting procedures.

New Officers of Illuminating Society

Gilbert K. Hardacre, manager of commercial sales of the Public Service Company of Northern Illinois, has been elected president of the Illuminating Engineering Society. Other new officers are: general secretary, Walter Sturrock; treasurer, Charles H. Goddard; vice president, Lee E. Taylor; directors, Allan E. Parker and P. M. Rutherford, Jr.

Meeting of Canadian Electron Microscopists

From June 12 to 14 representatives of the four Canadian electron microscope laboratories (University of Toronto, McGill University, National Research Council, and Shawinigan Chemicals, Ltd.) met with the research staff of Shawinigan Chemicals, Ltd., at Shawinigan Falls for presentation of papers concerning their own work and discussion of techniques and problems.

The Bulletin of Mathematical Biophysics

SEPTEMBER 1946

Some Considerations on the Structure of the Quadruped Extremities—N. RASHEVSKY

Tree Trunks and Branches as Optimum Mechanical Supports of the Crown: II. The Branches—MARTINUS H. M. ESSER

The Probabilistic Approach to the Effects of Radiations and Variability of Sensitivity.—I. OPATOWSKI

On the Relationship Between Response Time and Dosage of a Drug as a Function of its Mode of Entry—H. D. LANDAHL

The University of Chicago Press, Chicago, Illinois, Vol. 8, No. 3.

Joint Meeting of EMSA and ASXRED

On December 5, 6, and 7, 1946, a joint meeting of the Electron Microscope Society of America and the American Society for X-Ray and Electron Diffraction will occur at Mellon Institute and the University of Pittsburgh. All interested persons are welcome to attend. The program is as follows:

Thursday, December 5, 9:30 a.m. (Joint session)

1. Trends in x-ray diffraction work. DAVID HARKER, *General Electric Company.*
2. Trends in the use of the electron microscope in the study of fine structure. JAMES HILLIER, *Radio Corporation of America.*
3. The application of electron diffraction methods in the study of fine structure. ROBERT HEIDENREICH, *Bell Telephone Laboratories.*

2:00 p.m. (EMSA)

1. First-order theory of the three electrode electron gun with plane electrodes. S. G. ELLIS, *University of Toronto.*
2. Some observations on the shadow cast formvar replica technique. BEATRICE M. DEACON, S. G. ELLIS, AND W. G. CROSS, *University of Toronto.*
3. Electron microscopy of colloid systems. JOHN TURKEVICH, *Princeton University,* AND JAMES HILLIER, *Radio Corporation of America.*
4. Electron microscope examination of the microphysical properties of the polymer cuprene. JOHN H. L. WATSON, *Shawinigan Chemicals Limited.*
5. An effect of electron bombardment upon carbon black. JOHN H. L. WATSON, *Shawinigan Chemicals Limited.*
6. The electron microscope examination of greases. S. G. ELLIS, *University of Toronto.*
7. Artifacts produced by poor resolution. L. H. WILLISFORD, *Goodyear Tire and Rubber Company.*

2:00 p.m. (ASXRED)

1. Diffraction patterns of some non-crystalline soap phases. H. NORDSIECK, F. B. ROSEVEAR, AND R. H. FERGUSON, *Procter and Gamble Company.*
2. Triple chain length structure in mixed saturated triglycerides. E. S. LUTTON, *Procter and Gamble Company.*
3. Preparation and diffraction studies of synthetic diacid diglycerides. B. F. DAUBERT AND S. S. SIDHU, *University of Pittsburgh.*
4. The crystal structure of the sugars. R. PEPINSKY, *Alabama Polytechnic Institute.*
5. Business meeting.

6:30 p.m.

Joint dinner and address by FRANCIS O. SCHMITT, *Massachusetts Institute of Technology.*

Friday, December 6, 9:00 a.m. (EMSA)

Symposium on Biological Problems

1. Remarks on the borderline of physics, chemistry, and biology. L. MARTON, *Stanford University.*
2. Applications of the phase microscope in electron microscopy. OSCAR W. RICHARDS, *American Optical Company.*
3. Preparation of tissue for high speed sectioning. MARY C. SCHUSTER AND CLIFFORD E. GRAY, *Interchemical Corporation.*
4. Electron microscope studies on the structure of chlorophyll "crystals," "crystalline alpha carotene," "crystalline beta carotene," and "crystalline vitamin A." EDITH A. ROBERTS, *Vassar College.*
5. Further electron microscope studies of the structure of starch grain. MILDRED D. SOUTHWICK, *Vassar College.*
6. The mounting of bacteria for electron microscope examination. JAMES HILLIER AND R. F. BAKER, *Radio Corporation of America.*
7. A study with the high-voltage electron microscope of the structure and germination of the endospores of *Bacillus mycoides.* GEORGE KNAYSI, R. F. BAKER, AND JAMES HILLIER, *Radio Corporation of America.*
8. Demonstration with the high-voltage electron microscope of nucleus-like bodies in young cells of *Bacillus mycoides* grown in nitrogen free medium. GEORGE KNAYSI AND R. F. BAKER, *Radio Corporation of America*
9. Observations on actin and myosin extracted from muscle. M. A. JAKUS AND C. E. HALL, *Massachusetts Institute of Technology.*

9:00 a.m. (ASXRED)

1. An application of Fourier transforms to crystal structure analysis. DOROTHY WRINCH, *Smith College.*
2. Fourier synthesis of lattice disorders. W. H. ZACHARIASEN, *University of Chicago.*
3. Isothermal order—disorder in a salt system. ALEXANDER GRENALL, *Union Oil Company of California.*
4. The indexing of powder diffraction data. F. W. MATTHEWS, *Canadian Industries Limited.*
5. Ambiguities in diffraction analysis. A. L. PATTERSON, *Bryn Mawr College.*
6. A study of heterogeneous reactions by the electron diffraction and electron microscope techniques. EARL A. GULBRANSEN, *Westinghouse Research Laboratories.*

2:00 p.m. (EMSA)

Symposium on Biological Problems

1. Size distribution of tobacco mosaic virus in the early stages of infection. GERALD OSTER, *The Rockefeller Institute for Medical Research.*
2. Studies on the Newcastle disease virus (California strain). A. R. TAYLOR, *Duke University Medical School.*
3. An electron microscope study of the structures of some virus protein crystals. R. W. G. WYCKOFF, *National Institute of Health.*
4. Imminochemical aspects of sonic treated virus. SAUL MALKIEL, *The Rockefeller Institute for Medical Research.*
5. Business meeting.

2:00 p.m. (ASXRED)

1. Studies of a montmorillonite cracking catalyst. ALEXANDER GRENALL, *Union Oil Company of California.*
2. An x-ray diffraction study of a new cobalt carbide. L. J. E. HOFER AND W. C. PEEBLES, *U. S. Bureau of Mines.*
3. An x-ray diffraction study of chrome steel slags. G. P. CHATTERJI AND S. S. SIDHU, *University of Pittsburgh.*
4. Determination of silica in industrial dusts by means of the Geiger counter x-ray spectrometer. HAROLD P. KLUG, LEROY E. ALEXANDER, AND ELIZABETH BOLLMAN, *Mellon Institute.*
5. An x-ray diffraction study of the silver-magnesium alloy system. HAROLD R. LETNER AND S. S. SIDHU, *University of Pittsburgh.*
6. A long spacing x-ray diffraction camera. ORVIL BOLDUAN, *Massachusetts Institute of Technology.*

8:30 p.m. (Joint session)

Report on London Meeting
DAVID HARKER, *General Electric Research Laboratory*
I. FANKUCHEN, *Polytechnic Institute of Brooklyn*

Saturday, December 7, 9:00 a.m. (Joint session)

1. Specimen preparation for electron micrographs and diffraction patterns. LELAND L. ANTES, *University of Texas.*
2. Alterations and additional equipment for an R.C.A. console electron microscope. HENRY C. FROULE, *Armour Research Foundation.*
3. Spherical aberration of compound magnetic lenses. L. MARTON AND K. BOL, *Stanford University.*
4. Further studies on the magnetic electron microscope objective. JAMES HILLIER AND E. G. RAMBERG, *Radio Corporation of America.*
5. The bias focused electron gun. JOHN H. REISNER, *Radio Corporation of America.*
6. Photographic plates for use in electron microscopy and electron diffraction. ROBERT G. PICARD AND W. F. SWANN, *Radio Corporation of America.*
7. Calculated magnitude of Fresnel diffraction in shadow casting. C. E. HALL, *Massachusetts Institute of Technology.*
8. A review of some recent advances in electron microscopy microtomy. E. FULLAM, *General Electric Research Laboratories.*
9. Round table discussion of cross sectioning for the electron microscope. Invited speakers and discussion from the floor.

Ohmite Laboratory Dedicated

The new Ohmite Laboratory was dedicated in June in an all-day ceremony at the Armour Research Foundation of Illinois Institute of Technology to serve industry of the Middle West with precision electrical measurements established on the basis of and approved by the Bureau of Standards.

Oak Ridge Groups Merge

The Association of Oak Ridge Scientists and the Association of Oak Ridge Engineers and Scientists have been dissolved in favor of a single organization, The Association of Oak Ridge Engineers and Scientists. The amalgamation became effective at a general meeting of the membership of both organizations held in June. For the present, correspondence to the new organization or any officer or member thereof may be addressed to either P. O. Box 382 or P. O. Box 1251. All debts, assets, and activities of the two older organizations have been transferred to the new group and its officers.

New Physics Division at Cornell

One of the nation's first schools of engineering physics with a five-year curriculum was opened at Cornell University in September, "to train students in a combination of engineering principles, applied physics and mathematics for careers in research and development." According to Dean S. C. Hollister of the College of Engineering through which the new department is being administered, "The Department of Engineering Physics will bridge the area lying between the training in basic science and in fundamental engineering." He pointed out that the traditional curriculum in engineering falls short on the side of basic science, and the physics curriculum falls short on engineering principles.

The degree of Bachelor of Engineering Physics will be awarded upon completion of the five-year curriculum, and graduate work will be offered through the Graduate School of the University. Dr. Lloyd P. Smith is serving as director of the new department in addition to his post of chairman of the Department of Physics.

Training in Electron Microscopy at Toronto

Professor E. F. Burton of the McLennan Laboratory Department of Physics, University of Toronto, has announced that a graduate course will be given next year for the purpose of training graduate students in all phases of electron microscopy theory and practice. The course next year will be limited to six students who must have had advanced training in mathematics and physics. The course will be accepted as work toward the degree of M.A., and candidates must be acceptable to the School of Graduate Studies, University of Toronto.

New Booklets

Split Ballbearing Corporation, Lebanon, New Hampshire, has issued Catalog No. 84, giving full specifications on sizes, load ratings, etc., on a complete line of divisible race, ball, roller, and thrust bearings. 30 pages. Free on request.

United Electric Controls Company, 69 A Street, Boston 27, Massachusetts, has published a new catalog giving information and engineering details on all their thermostats and pressure switches. 30 pages. Free on request.

H-B Instrument Company, 2518 North Broad Street, Philadelphia 32, Pennsylvania, announces its Catalog No. 14 listing temperature controls, relays, and thermometers. 16 pages. Free on request.

The "Milvay Notebook" for Spring, 1946, published by Chicago Apparatus Company, Chicago 22, Illinois, has the following table of contents:

Titrimetric Procedures.
Microscope Camera Attachment.
Spectrophotometer.
Synthetics and More Synthetics.
Glas-Col Flask Heaters.

34 pages, numerous illustrations.

The Gaertner Scientific Corporation, 1201 Wrightwood Avenue, Chicago 14, Illinois, recently issued a leaflet describing the L123 precision spectrometer with two microscopes, a camera attachment, polarizing attachment, Babinet and Babinet Soleil compensators, Abbe-Lamont auto-collimating eyepiece, filar eyepiece micrometer, and various other items. 8 pages. Free on request.

Eastman Kodak Company, Rochester 4, New York, has announced a third revised edition of the *Kodak Data Book on Formulas and Processing*. Containing a comprehensive list of Kodak formulas together with a discussion of the principles and procedures for processing films, plates, and papers, the new edition incorporates a number of changes to bring it up to date.

Bulletin 155–64 of The Gaertner Scientific Corporation, 1201 Wrightwood Avenue, Chicago 14, Illinois, describes the new Gaertner large toolmaker microscope. 8 pages. Free if request is made on company letterhead.

General Electric Company's *Special Products Digest* for June 1946 features an article on the G-E 100-million-volt betatron and its use in atomic research. 4 pages.

The July-August issue of *Westinghouse Newsfront* contains a description of the newest methods of cleaning and drying indoor air, and a page of pictures of different uses of the servo-mechanism. 4 pages.